

Agrobiological Sciences Rice Full-Length cDNA Project Team,
 Kikuchi, S., Sato, K., Nagata, T., Kawagashira, N., Doi, K.,
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
 Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group, Otsu, Y., Murakami, K.,
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
 Kurowski, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
 Narioka, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
 Kawaguchi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., KIKEN:
 Kawai, J., Carninci, P., Adachi, O., Aizawa, K., Arakawa, T., Fukuda, S.,
 Hara, A., Hashidume, W., Hayatsu, N., Inotani, K., Ishii, Y., Itoh, M.,
 Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
 Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Yoshino, M. and Havareshizaki, Y.

575 ATGCTGGACCTTTCAGACACTACTGTGGTGCATTTGAGCGCTGCATATACCCAGGCAACTCGCGAA 634
543 GAGCACATCATTTATCTGAAACAGCAACTTGTGGAGCTGTGGCTCGAGCATTCATTAACATTTTGG 602
635 GCCATCTCTCTCTGTGAGCACAGCTTGTGTGACTGTGGTTTGTCAATTCACAAATTTCCG 694
603 ATGCAATGGAGGTTTGGCTTCCCAAGCCTTTTGATATAGTTAAGTACAAATCGAGGCAATCGA 662
695 ATGCAAGGAGGCTTCCATCCAGGCTTTGAAATACATCAAAATACAAATGGTGGCTTGA 754
663 CACAGAACAGACTTATCCATACCTTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGAA 722
755 CACTGAGGAATCTTACCCCTTACCAAGGTGTCAATGGAATCTCCAAGTTTAAAGAATGAGAA 814
723 TGTGTGTGTCAGGTCATTTGATTCGATTAACATCACCTCGGCTGTGAGGATCAGTTGAA 782
815 TGTGTGAGTCAAGGTTTGGACTCGGTTTAAACATCACCTGGTGTCTGAGGATGAATGAA 874
783 GCATGCAAGTGGCTTGTGTCGTCCTGAGTTTGGCTTGCATTTGAGTTGTGAAAGTTTCAA 842
875 GGAATGCTGTTGGTGTGGTTCGCCCAAGTTAGTTTGGCTTGCAGGTGATCACTGGTTTCAG 934
843 TCTGTACAGAAAGGTGTATACAGCAGTGACACTGTGGAAGAGATCCATGATGATGAA 902
935 GCTGTACAGAGCGGAGTTTACATGAGCAGCAATGTGGAATCTACACCGATGATGAA 994
903 CCACGAGTTTTCGCGTCGTTATGAGTTCGAGGACGGGATTCCTTTATTTGGCTCATCAA 962
995 CCACTGCTTTCGCTGTTGGCTACGGTGTGGAAGATGGTGTACCTTACTGGCTCATCAA 1054
963 GAACTCATGGGTACAAATTTGGGTGACAAATGGCTTACTTTAAGATGAACTCGGCAAGAA 1022
1055 GAATCATGGGGCTGACTGGGGTATGAGGGTACTTCAAGATGAAATGGCAGAA 1114
1023 CATGTGTGTTGGTGCAACTTGGCGCATCTTATCCCATTTGGGC 1064
1115 CATGTGGGTGTTGCTACGTGTGCATCTTACCCTATTGTGCG 1156

RESULT 4
MZECYPA
LOCUS MZECYPA
DEFINITION Corn mRNA for cysteine proteinase, clone CCP2, complete cds.
ACCESSION D45403
VERSION D45403.1 GI:1688044
KEYWORDS cysteine proteinase.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1382)
Domoto,C., Watanabe,H., Abe,M., Abe,K. and Arai,S.
Isolation and characterization of two distinct cDNA clones encoding corn seed cysteine proteinases
Biochim. Biophys. Acta 1263 (3), 241-244 (1995)
JOURNAL 96004895
MEDLINE 7548211
REFERENCE 2 (bases 1 to 1382)
Domoto,C.
Direct Submission
AUTHORS Submitted (31-JAN-1995) Chieko Domoto, Atomi Junior College;
Ohtsuka 1-5-2, Bunkyo-ku, Tokyo 112, Japan
JOURNAL (Tel:03-3941-8161(ex.575), Fax:03-3945-1836)
On Nov 27, 1996 this sequence version replaced gi:644489.
COMMENT Sequence location (26-Oct-1996) by: Chieko Domoto.
FEATURES
source
1..1382
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone="CCP2"

[illegible]

191 GGAAGAGCTACGGATCGGAGGAGAGATCAAGAGAGGTTCCGGATCTTCGTGAGAAATC 250
232 GCAAGAGCTACGGAGCGCGGAGAGTGAAGAGCGGTTCCGGATCTTCGTGAGAGCC 291
251 TAGCGTTATCCGCTCCACCTAATCGGAGGATCTGCTGATACCTTAGGAATCAACCAAT 310
292 TCGAGAGAGTCCGCTCCACCAACCGGAGGCGCTCTCTACAGCTCGGATCAACCGAT 351
311 TCGCGACCTGACCTCGGAGGAAATCCGGAGCAATCGCCTTGTTGGCGCGCAGAACTGCT 370
352 TCTCTGACATGACCTGGGAGGAGTCCAGGCAACCAAGCTTGGCGCGCGCAGACCTGCT 411
371 CGGCGACTGCGCATGGAACCAACCGGTT---TGTGATGCGGCTTCTCTGTAAACGAGG 427
412 CGGCGACGCTCGCGGCAACCACTGATGCGGAGCGCAACGCGCTCCCGGAGCAACAAAG 471
428 ATTGAGGAGCAAGGAGATGAGCGCTGTAAAGGACCAAGGAAGCTGTGGATCTTGTCT 487
472 ACTGGAGGAGACTGGGATCGTTAGCCCTGTGAAGACCGGCAAGCTGCGGCTCTGCT 531
488 GGACTTTCACTACTGAGGACACTAGAGGCTGCATATACAGCTAACTGGAAGAGCA 547
532 GGACCTTCAGCACTACTGAGCACTTGAAGGAGCATATACTCAGGCCACTGGAAGAAACA 591
548 CATCAATATCGAAGCACTTGTGGACTGTGCTCAGCAATCAATAAATTTGGATGCA 607
592 TCTCTCTTCTGAGCAACAGCTGTTGACTGTGCGGTGCATACAAATTTGGATGCA 651
608 ATGAGGTTTCCCTTCCCAAGCCTTTGAATACGTTAAGTAACTAGGAGGATGAGCAAG 667
652 ACGGAGGCTTCCATCCCAAGCACTTGAATACAGTCAATGAGTGGCACTGACACCG 711
668 AACAGACTATCCATACCTTGGTGTCAATGCTATCGCACTTCAAGCAGGAGATCTG 727
712 AGGAGTCTTACCCCTTACAAGGCTGTCAATGGGCTGTGCAATACAGGCTTGAATGCTG 771
728 GTGTCAAGGTCATTGATTCGATAAACAATCAACCTCGGTGCTGAGGATGAGTTGAAGCATG 787
772 CAGTTACAGTTGCGGACTCGGTCAACATCACTCTGAAATGCTGAGGATGAACTGAAGAATG 831
788 CAGTGGGCTTGGTGGCTCAGTTAGCTTGCATTTGAGTTGTGAAGGTTTCAATCTGT 847
832 CTGTGGGTTGGTTGCGCCAGTCAAGTGTGCTTGAAGTGTGATCGAGGTTTCAAGCAGT 891
848 ACAAGAAAGGTGTATACAGCAGTACACCTCTGGAAGAGATCCCAATGATGGAACCAAG 907
892 ACAAGAGCGGAGTTTACACTAGTGACCATTTGTGGAATCTACTCTGACGATGTCAACCAAG 951
908 CAGTTCTTGGCTGGTTATGAGTFCGAGGACGGGATTCCTTATTTGGCTCATCAAGAACT 967
952 CTGTTCTAGCTGTTGGCTACGCTGCGGAGATGCGCTCCCGTATTTGGCTCATCAAGAACT 1011
968 CATGGGTTACAAATTTGGGCTGACATGCTACTTTAAGATGGAATCGGCAAGAACTATG 1027
1012 CATGGGCGCGACTGGGCTGAAGAGGTTTACTTCAAGATGGAATGCGCAAGAACTATG 1071
1028 GTGTGTTTGAACCTTGGCACTCTTATCCCATTTGGCTG 1066
1072 GTGTGCTGCTACCTGCGGCTCTACCTATCTTGGCGG 1110

RESULT 7
AR231148
LOCUS AR231148 1407 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 107 from patent US 6451604.
ACCESSION AR231148
VERSION AR231148.1 GI:27271935
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1407)

AUTHORS Plinn, B. and Lasham, A.
TITLE Compositions affecting programmed cell death and their use in the modification of forestry plant development
JOURNAL Patent: US 6451604-A 107 17-SBP-2002;
FEATURES Location/Qualifiers
source 1. 1407
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 50.6%; Score 541.4; DB 6; Length 1407;
Best Local Similarity 70.5%; Pred. No. 4.2e-138;
Matches 738; Conservative 0; Mismatches 306; Indels 3; Gaps 1;
21 CTTCTACTCTTCTCTCGCCCTCTCGCTCGGCTCGCGGTTTCGAGAGAACAA 80
Db 87 CGTCTCTCTCTCTCGCGCTCTCGCTCTCCCGCGGCTCGAGCTTCGAGAGTCCAA 146
21 TCCATATCCCGTCCGTTTACACA---AAGGCTTACATCGATTTGAGCTTCCTCGGCT 137
Db 147 CCCCATCCGCTCTTCCCGACGCGGCTCTCGGACCTCGAGTCTCTCCATCGTCCAGAT 206
21 CTTGGAGTTCGCGCAGCGCTTCACTTCCGACGCTTCGCGCGGCTTCGAGAGTCCAA 197
Db 207 CGTGGCGCACCCCGCACGCTTCTCTTCCCGCGCTTCGCCAACAGGATATGAGAGAG 266
21 CTACGAGTCCGAGGAGGATCAAGAGAGGTTCCGGATCTTCTGTTGGAATCTTAGCGTT 257
Db 267 GTACGAGACCGGAGGAGATCAAGCTCGGTTTCAGATCTTCAGGAGAGATCTCAAGTT 326
21 TATCCGCTCCATATTCGGAAGATCTGTCTATACCTTAGGAATCAACCAATTCGCCGA 317
Db 327 GATCCGATCCACCAACAGAGAGGCTTGCCTTACACCTCGGTGTCAATTAAGTTTGTCA 386
21 CCTGACCTCGGAGGATTCGCGACCAATCGCTTGGTGGCGCGCAGAACTGCTCGCGAC 377
Db 387 TTGGAGCTGGAGAGGTTTCAAGAGGCAAGATCTGGAGTGTCTCAAAATGCTCTGCCAC 446
21 TGCGCATGGAACCAACCGGTTTGTGATGCGCTGCTTCTGTAACGAGGATTTGAGGGA 437
Db 447 CACCAAGGCAACCAACGCTCACCGAAGCTCTTCCCGAGATGAAAGACTGGAGAGA 506
21 GCAAGGATAGTGAAGCGCTGTAAAGACCAAGAGCTGTGGATCTTGTGAGCTTTTCAG 497
Db 507 AAAGGCAATGTAAGCGCCAAATTAAGATCAGGGGCACTGTGGATCTTGTGAGCTTTCA 566
21 TACTACTGGAGCACTAGAGCTGCATATACAGACTTAACAGAGGAGGACATCATATC 498
Db 567 TACCCTGGAGCTTTGAGGCTGCTTATCAACAGCAATTCGGGAAACAAATCTCTCTGTC 626
21 TGAACAGCAACTTGTGGATGCTGCTCAGCATTAATTAATTTGGATGCAATGGAGTTT 617
Db 627 TGAGCAGCAGCTCGTGGACTGTCTGGGCTTTTCAACACTTTGGATGTAGTGGTGGACT 686
21 GCCTTCCCAAGCCTTTGAATAGCTTAAAGTACAAATGAGGATCGACACAGAAACAGACTTA 677
Db 687 GCAATCCCAAGCCTTTGATGATCGTCAAGTACAAAGTGGCTTGTATCCGAGGAGCAATA 746
21 TCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGGT 737
Db 747 TCCTTATACCGCAGTGGATGGTGTAGCTGCAAAATTCGCGGTGATTAATGTTGGTGTCCAAGT 806
21 CATTGATTCGATAAACAATCACCTCGGCTGTGAGGATGAGTTGAAGCATGCAAGTGGGCTT 797
Db 807 GCTCGACTCTGTAAATCACTACCTTGGGCTGTGAGGATGAACCTTAAGCATGCAAGTGGCTT 866
21 GGTGCGTCCAGTTAGGCTTGCATTTGAGGTTGTGAAGGTTTCAATCTGTACAGAAAGG 857
Db 867 CGTCCGCGCAGTGTGAGTGTGCAATTCAGGTTCCAGGTTGTGAAGAGCTTCAGATTTGACAAAGTCGG 926
21 TGTATACAGAGTGAACCTGTGGAAGAGATCCAAATGATGTGAACCAAGCAGTCTTTC 917
Db 927 TGTCTACAGAGCGATACATGCGGTAGCACTTCCATGGATGTGAACCATGCTCTTCTCGC 986

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 29, 2004, 17:17:07 ; Search time 2963 Seconds
(without alignments)
15666.686 Million cell updates/sec
Title: US-10-087-714-1
Perfect score: 1071
Sequence: 1 atggcagctaagctctctt.....atcccatgtggtgtgtgag 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sv.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Length	DB ID	Description
1	576.4	53.8	1486	8 AK066748 Oryza sat
2	569.6	53.2	1412	8 ZMSBE1
3	569.6	53.2	1442	6 A43549
4	568	53.0	1382	8 MZECYPA
5	558.8	52.2	1468	8 RICOZC
6	551.8	51.5	1345	8 LMU249847
7	541.4	50.6	1407	6 AR231148
8	518.8	48.4	1332	8 AF454960
9	517.2	48.3	1444	8 PAU93166
10	511.2	47.7	1340	8 AB032168
11	488.4	45.6	1305	8 AF233883
12	488.4	45.6	1352	8 AY088662
13	488.4	45.6	1355	8 AF083703
14	486.8	45.5	1331	8 BT000676
15	486.8	45.5	1370	8 AF360273
16	486.2	45.4	1347	8 PHU31094
17	485.4	45.3	1388	6 A43538
18	485.4	45.3	1366	8 LECYP3GN
19	485.2	45.3	1366	8 BT000674
20	485.2	45.3	1409	8 BT000673
21	474	44.3	1381	8 PSRNACP
22	473.6	44.2	1484	8 CACG2
23	472.4	44.1	1361	8 FSA278699
24	467.6	43.7	1402	8 AY091771
25	344	32.1	937	8 AY383240
26	304	28.4	536	8 AF213939
27	280.2	26.2	677	6 AR231152
28	259.8	25.2	1362	10 RNCATH
29	267	24.9	1360	10 RNCATH
30	262.6	24.5	1372	10 BC006878
31	261.6	24.4	1149	4 SSAFE001169
32	261	24.4	1381	10 MMCTSH
33	247.2	23.1	972	9 AF426248
34	247.2	23.1	1008	9 AF426247
35	247.2	23.1	1399	9 HSCATH
36	245.6	22.9	1470	9 BC002479
37	244	22.8	1438	9 AK130158
38	238.4	22.3	1106	6 AX774741
39	238.4	22.3	1106	9 HSCATH
40	205.2	19.2	475	6 AR231161
41	200.8	18.7	841	8 AF242733
42	197.6	18.5	535	8 NTPPCYSPT
43	182.8	17.1	498	10 MMU18464
44	181.6	17.0	1168	3 AY336797
45	181.2	16.9	498	9 HOS18461

ALIGNMENTS

RESULT 1
AK066748
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013074D19, full insert sequence.
ACCESSION AK066748
VERSION AK066748.1 GI:32976766
KEYWORDS FLI CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-length cDNA Consortium, National Institute of

AK066748 1486 bp mRNA linear PLN 24-JUL-2003
Oryza sativa (japonica cultivar-group) cDNA clone:J013074D19, full

QY 918 CGTCGGTTATGGAGTCGAGGACGGGATTCCTTATTTGGCTATCAAGAACTCATGGGTAC 977
 DB 987 TGTGGTTATGGAGTTGAAGATGGTGTTCCTGCTCATCAAGAAATTCCTGGGAGC 1046
 QY 978 AAATTGGGGTGACAACTGCTACTTTAGATCGAACTCGGCAAGAACATGTGTGTGTC 1037
 DB 1047 AGACTGGGGTGACCAACGATACCTTCAAGATGAGATGGGAAGAACATGTGTGGATCGC 1106
 QY 1038 AACTTGGGCATCTTATCCCATGTGGC 1064
 DB 1107 TACTTGTGCATCATACCTGTTGTGGC 1133

RESULT 8
 AF454960 1332 bp mRNA linear PLN 14-JAN-2002
 LOCUS
 DEFINITION Brassica oleracea senescence-associated cysteine protease (CPs)
 mRNA, complete cds.

ACCESSION AF454960
 VERSION AF454960.1 GI:19141288

KEYWORDS
 SOURCE
 ORGANISM

Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
 1 (bases 1 to 1332)
 Coupe, S.A., Sinclair, B.K., Watson, L.M. and Eason, J.R.
 Cysteine proteases and broccoli senescence: cloning,
 characterization and contribution to the process
 Unpublished

JOURNAL
 2 (bases 1 to 1332)
 Coupe, S.A., Sinclair, B.K., Watson, L.M. and Eason, J.R.
 Direct Submission
 Submitted (05-DEC-2001) Crop and Food Research, Private Bag 11 600,
 Palmerston North 5301, New Zealand

FEATURES
 Location/Qualifiers
 1..1332
 /organism="Brassica oleracea"
 /mol_type="mRNA"
 /db_xref="taxon:3712"
 1..1332
 /gene="CP5"
 32..1111
 /gene="CP5"
 /note="BoCP5; similar to Arabidopsis aleurain-like
 protease; expression increases as the broccoli heads
 senesce after harvest and during storage"
 /codon_start=1
 /product="senescence-associated cysteine protease"
 /protein_id="AAL60582.1"
 /translation="MMSVRIILPSSAVLLILAASTAESIGFDSNPIMVSDRLREVE
 ESVVILGDSRHVIFSFARFAHRYGRYENAEMLKRFIFKENLDLIRSTNKGILSYK
 LVNQPDMQWFOFTKLGAQNCATLKGHTLGTALPETKMDRDTGIVSPVDQ
 GCGCWTFSTTGALAEAYHQAFKGISLSEQQLVDCGAFNNYCGNGGLPSQAEVI
 KNGGLDTEAPYPTGEQDTKYSAEENVGEVLDVSNITLGAEBELKHAUGLVRFVSI
 AFEVTHSLFKSGYSPSHCGQTPMDNVHVLAVGVGIEDGVPPYLLIKNSWGADWDG
 KGYFRMEMKNNMGICATCASYEVA"

gene
 CDS

Query Match 48.4%; Score 518.8; DB 8; Length 1332;
 Best Local Similarity 68.6%; Pred. No. 7,2e-132; Indels 0; Gaps 0;
 Matches 715; Conservative 0; Mismatches 327;

QY 23 TCCTACTCTCTCTGCTCCGCCCTCTCCGTCGCGCTCGCCGGTTTCGAAGAAGACAATC 82
 DB 65 TTCTGCTGATCTCTCATCGCGCATCGAAGCGGAGAGTATCGGATTTGATGATCAAACC 125

QY 83 CAATCGGTCGGTTACAAAGCCCTGACTCGATTGAGCCTCCATCCTCGCGCTCTTG 142
 DB 126 CGATCCGTTATGGTCTCCGACCGGTCTCCGGAGGTAGGAATCCCGTTGCCAGATCTTAG 185

ORIGIN
 Query Match 48.4%; Score 518.8; DB 8; Length 1332;
 Best Local Similarity 68.6%; Pred. No. 7,2e-132; Indels 0; Gaps 0;
 Matches 715; Conservative 0; Mismatches 327;

QY 23 TCCTACTCTCTCTGCTCCGCCCTCTCCGTCGCGCTCGCCGGTTTCGAAGAAGACAATC 82
 DB 65 TTCTGCTGATCTCTCATCGCGCATCGAAGCGGAGAGTATCGGATTTGATGATCAAACC 125

QY 83 CAATCGGTCGGTTACAAAGCCCTGACTCGATTGAGCCTCCATCCTCGCGCTCTTG 142
 DB 126 CGATCCGTTATGGTCTCCGACCGGTCTCCGGAGGTAGGAATCCCGTTGCCAGATCTTAG 185

RESULT 9
 PAU93166 1444 bp mRNA linear PLN 21-JAN-1998
 LOCUS
 DEFINITION Prunus armeniaca cysteine protease mRNA, complete cds.
 ACCESSION U93166
 VERSION U93166.1 GI:2677827
 KEYWORDS

QY 143 GCAGTTGGCGCCACGCTTCCACTTTCGACCGCTTCGCGCAGGTACGGAAGAGCTACG 202
 DB 186 GCGAATCAGCCACCGTATCTCTTCCTGCTCGCTTCGCTCACAGGTACGGAAGAGGTACG 245
 QY 203 GATCGGAGGAGGAGATCAAGAAAGAGTTTGGGATCTTCGTGGAGAACTTAGCGTTTATCC 262
 DB 246 AGAACGCGGAGGAGATGAAGCTCAGTCTCGATATCAAGGAGAACTTGGATTTCATCA 305
 QY 263 GGTCCACTAATCGGAAGGATCTGTGTATACCTTAGGAATCAACCAATTCGCGGACCTGA 322
 DB 306 GATCCACCAATAAGAAAGGCTTGTCTTCAAACTCGGTGTCATCAGTTTGTGATATGA 365
 QY 323 CTTGGGAGGAATTCGCGACCAATCGCTTGTGTCGCGCGCAGAACTGCTCGGCGACTCGC 382
 DB 366 CGTGGCAAGAGTTTCAAAGGACCAAGCTTGTGTGCTCAAAACTGCTCTGCCACTTTAA 425
 QY 383 ATGGAACACCGGTTTGTGATGGGTGCTTCTGTACAGAGGATTTGGAGGGGCGAAG 442
 DB 426 AGGGAACCCACAACTCACAGGAGAGCTCTTCCGGAACCAAGACTCGGAGAGATG 485
 QY 443 GGATAGTGAGCCCTGTAAAGGACCAAGGAAGCTGTGGATCTTGTGGACTTTTCAGTACTA 502
 DB 486 GTATTGTTAGCCCTGTCAAAGACCAGGGGGTGTGGATCGTCTGGACATTCAGTACTA 545
 QY 503 CTGGAGCACTAGAGGCTGCATATACACACTAACTGGAAGAGACATCATTTATCTGAAC 562
 DB 546 CTGGAGCGCTTGAGGCGCTTATCACCAAGCATTTTGGAAAAGGAATATCACTCTCTGAGC 605
 QY 563 AGCAACTTGTGACTGTGCTCAGCATTCATTAACATTTTGGATGCAATGAGGTTTGCCTT 622
 DB 606 AACAGCTTGTGATTTGTGTGGAGCTTTCACAACTATGCTGCAATGCGTGGCTTCCCTT 665
 QY 623 CCCAAGCCTTTGAATACGTTAAGTACAATGGAGGATTCGACACAGAACAGACTTATCCAT 682
 DB 666 CCCAAGCCTTTGAATACATCAATCCAAAGCGGCGCTCGACACGGAAGCTTTATCCTT 725
 QY 683 ACCTTGGTGCATGTTCTGCAACTTCAAGCAGGAGAACTGTTGTGTCAAGGTCTATTG 742
 DB 726 ACACCGGTGAAGATGGAACCTGCAAGTATTCCTGCAAAACGTCGGTGTAGAGTCTTGG 785
 QY 743 ATTTCGATAAACATCACCCCTGGGTGCTGAGGATGAGTTGAAGCATGAGTGGGTTGGTGC 802
 DB 786 ACTCAGTCAACATTAATCTCTGGGTGCTGAAGATGAAGTGAAGCATGCGGTTGGATTGTAC 845
 QY 803 GTCCAGTTAGCTGCTATTGAGTTGTGAAAGGTTTCAATCTGTACAAAGAGGTGAT 862
 DB 846 GACCAGTAAGCATAGCATTTGAGGTTTATACACTGTTCCGGCTTTTACAGAGCGGAGTTT 905
 QY 863 ACAGCAGTGACACTGTGGAAGAGATCCCAATGATGTGAACCAACCGCAGTTCTTGGCGTCG 922
 DB 906 ACAGCGNATGCTACTGTGGACAGACTCCCAATGCAAGCAACCGCTGTTTGGCGGTTG 965
 QY 923 GTTATGAGTTCGAGGAGCGGATTTCTTTATTTGGTCTCATCAAGAACTCATGGGTTACAAAT 982
 DB 966 GTTATGGAATTGAAGACGGTGTTCATATTTGGCTTTTAAAGAACTCATGGGAGCAGATT 1025
 QY 983 GGGGTGACATGCTACTTTAAGATGGAACCTCGGCAAGAACATGTGTGTGTGTGCAACTT 1042
 DB 1026 GGGGTGCAAAAGTTTCTTCAAGATGAGATGCGGAGAACATGTGTGTGTGTGCGACAT 1085
 QY 1043 GCGCATCTTATCCCATTTGTCG 1064
 DB 1086 GTGCATCGTACCCGGTTGTGGC 1107

RESULT 9
 PAU93166 1444 bp mRNA linear PLN 21-JAN-1998
 LOCUS
 DEFINITION Prunus armeniaca cysteine protease mRNA, complete cds.
 ACCESSION U93166
 VERSION U93166.1 GI:2677827
 KEYWORDS

602	CTATCTGAGCAGCAGCTTGTGGAATGCTGCTGGAGCTTTTAAATAACTTTGGCTTGCATGCT	661
613	GGTTTGGCTTCCCAAGCCCTTTGAATACGTTAAAGTACAAATGAGGATCGACACAGAAACAG	672
662	GGGCTCCCATCACAGACCTTTGAGTATATTAATCCATGGTGGTCTTGACACTGAAGAA	721
673	ACTTATCCATACCTTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCT	732
722	GCATATCCATACACTGCGCAAGATGGCTTATGTAAATTTCTCATCAGAAAAATGTTGGTGTCT	781
733	AAGTCATGATTTGATATAACATCACCTGGTGGTCTGAGGATGAGTTGAAGCATGCGAGTG	792
782	AAAGTCATCGAATTCGTCATATTAACCTGGTGGTCTGAAGATGACTAAATAACGCGGTT	841
793	GGCTTGGTGGCTCCAGTTAGCGTTGCAATTTGAGGTTGTGAAAGTTTCAATCTGTACAAG	852
842	GCAATGGTTAGGCCCGTTAGTATAGCTTTTGAGGTGATAAAAGGTTTCAACAATAACAAG	901
853	AAAGGTGATACAGCAGTGACACCTGTGGAAGAGATCCAAATGGATGTGAACCAACGCGATT	912
902	AGTGTGTTTACACACGACCCGAATGCGGCAACACTCCCATGGATGATTAACCAATGCTGTT	961
913	CTTGGCGCTGGTGTATGGAGTCGAGACGCGAATCTCTTATTTGGCTCATCAAGAACTCATGG	972
962	CTTGTCTGGGTGTACGGTGTGAAATGGTGTTCCTATTGGCTCATCAAGAATTCATGG	1021
973	GGTACAAATTTGGGTGACAAATGGCTACTTTAAGATGGAACACTCGGCAAGACATGTTGGT	1032
1022	CGAGCAGATTTGGGGTGCACAAATGGATCTTCAAAATGGAGATGGGAAAGAACATGTTGGT	1081
1033	GTTCGAACTTGGCGATCTTTATCCCATTTGGGC	1064
1082	ATTGCACTTGGCGATCTTACCTGTCGTGC	1113

RESULT 11

AF233883

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

AF233883

Arabidopsis thaliana AALP protein mRNA, complete cds.

AF233883.1

GI:7230639

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 1305)

Ahmed,S.U., Rojo,E., Kovaleva,V., Venkataraman,S., Dombrowski,J.E., Matsushita,K. and Raikhel,N.V.

The plant vacuolar sorting receptor ALEP is involved in transport of NH(2)-terminal propeptide-containing vacuolar proteins in Arabidopsis thaliana

J. Cell Biol. 149 (7), 1335-1344 (2000)

20330156

10871276

2 (bases 1 to 1305)

Raikel,N.

Direct Submission

Submitted (13-FEB-2000) MSU-DOE Plant Research Lab, MSU, Plant Biology Building, East Lansing, MI 48824, USA

Location/Qualifiers

1..1305

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

10..1086

/note="similar to aleurain"

/codon_start=1

/product="AALP protein"

/protein_id="AAF43041.1"

/db_xref="GI:7230640"

/translation="MSAKTILSSVVLVLAASAAANIGFDESPIRMVSDGLREVEE
SVSQILQSRHVLSPARFTHRYKKYQNVVEEMKLRFSIPKENLDLRSTNKKGLSYKL
GVNQFADLTWQEFQRTKLGAQNCATLKGSHVTEAALFETKDWREDSIVSPFRDQG
CGCSMTFTSTGALEAAHQAFKGLSLSEQQLVDCAGAFNNYNGNGLPSQAFPIK
SNGGLDTEKAYPTGKDETKFSAENVGVQLNSVNITLGADELKHLVGLVRPVSIA
FEVTHSPRLKSGVYTDSCGSGTSPMDVNHAVLAVGVGVEDVPIYMLIKNSWGDWGD
GYFQWENKMKMGIIATCASYPVVA"

ORIGIN

Query Match 45.6%; Score 488.4; DB 8; Length 1305;
Best Local Similarity 67.0%; Pred. No. 1.8e-123;
Matches 693; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY	32	TCCTGGTCTCGCCCTCTCGTCCGCTCGCGGTTTCGAAGAAGACAATCCAAATCCGGT	91
DB	50	TTCTCGTCGCGCATCAGCAGCGCTAATATCGATTCGATGAGTCAAAACCGATCCGAA	109
QY	92	CCGTTACACAAGCGCTGACTCGATTGAGCGTCGCCATCTCGGGGTCCTTGGCAGTTGCC	151
DB	110	TGGTCTCGATGCTCTCGGAGGTAGAAGAACTGTGTTCCAGATCTTAGTCAATCTC	169
QY	152	GCCAGCGCTTCCATCTCGCAGGTTCCCGCAGGTACGGGAAGAGCTACGGATCGGAGG	211
DB	170	GTCAGTCTCTCTTCTGCTCGCTTCACTACCGATATGTTAAAGTATCAGAACGTGG	229
QY	212	AGGAGATCAAGAAGAGGTTCCGGATCTTCGTGGAGAACTTAGCGTTTATCCGTTCCACTA	271
DB	230	AGGAGATGAACTTCGATTCTCGATTTTCAAGGAGAACTTGAATTCGATCAGATCCACCA	289
QY	272	ATCGGAGGATCTCTGTATACCTTAGGAATCAACCAATTCGGACCTCAGCTGGGAGG	331
DB	290	ACAAGAAAGGTTATCTTACAAATCGCGTGTAACTCAATTTGCTGATTCACATGGCAAG	349
QY	332	AATTCGCGACCAATCGCCTTGGTGGCGCGCAGAACTGCTCGGACCTGCGCATGGAAAC	391
DB	350	AGTTTCAAGACCAAGCTTGCTGCTCAGAACTGCTCGCCACTTTAAAGGGCAGCC	409
QY	392	ACCGGTTGTGATGGCGTGCTCTGTACAGGAGGATTCGAGGAGCAAGGATAGTGA	451
DB	410	ACAAGGTACAGAAGCAGCTCTTCTGAAACAAAGAACTGGAGAGAGATGGTATCGTTA	469
QY	452	GCCCTGTAAGAGCAACAGGAAGCTGTGATCTTCTGGAATTTTCAGTACTACTGGAGCAC	511
DB	470	GTCGGTCAAGATCAGGAGGTTGTGATCTTCTGGAATTTTCAGTACTACTGGAGCAC	529
QY	512	TAGAGGTGCATATACACAGTAACTGGAAGAGCAATCATATCTGAAACAGCAACTTG	571
DB	530	TTGAGGCAGCTTACCATCAGGCAATTCGAAAGGAATATCTCTCTGAGCAACAGCTTG	589
QY	572	TGGATGTGCTCAGCATTCATTAACCTTGGATGCAATGGAGGTTTCCTTCCCAAGCCT	631
DB	590	TGGATGTGCTGAGCTTTCATTAACCTTGGATGCAATGGAGGTTTCCTTCCCAAGCCT	649
QY	632	TTGAATAGTTTAAGTACAAATGAGGAGCTCGACAGAACAGACTTATCCATACCTGGTG	691
DB	650	TTGAATACATAAATCCAAAGGTCGCTCGACAGACAGAGAAAGCTTATCCCTATACCGGTA	709
QY	692	TCAATGTATCTGCAACTTCAAGCAGGAGATGTTGTTGTCAAGTCAATGATTCGATAA	751
DB	710	AAGATGAACCTGCAAAATTTTTCAGCTGAAACGTTGTGTGACAAAGTCTCAACTCAGTCA	769
QY	752	ACATCACCTCGGTGCTGAGGATGATGTAAGCAGTTCAGTGGGCTTGGTGGTCCAGTTA	811
DB	770	ACATTAATCTGGTGTGTAAGATGAACTGTAAGCATCGGTTGGATTTGATCGGCCAGTAA	829
QY	812	CGGTGCAATTTGAGTTGTGAAGGTTTCAATCTGTACAGAAAGGTTATACAGCAGTG	871
DB	830	GCATAGCATTTGAGGTTATACACTCGTTCCGGCTTTTCAAGAGTGGAGTTTACACTGATA	889
QY	872	ACACCTGGAAGAGATCCAAATGATGTAACCAAGCAGGTTCTTCCGCTCGGTTATGGAG	931
DB	890	GTCACCTGGAAGTACTCAATGATGTAAGCAACCAAGCGGTTTGGCCGTTGGTTATCGAG	949

QY	932	TCGAGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGCTGACA	991
DB	950	TTGAAGACGGTGTACCATATTGGCTTATTAAGAACTCATGGGACGCGATTGGGCGACA	1009
QY	992	ATGGCTACTTTAAGATGAACACTCGGCAAGACAATGTGGTGTTCGCAACTTGGCGCATCTT	1051
DB	1010	AAGGTTACTTCAAGATGAGATGGGGAAGACAATGTGGTATTGCTACATGTGCATCAT	1069
QY	1052	ATCCCATTTGGCT 1065	
DB	1070	ACCCGTTGTGGCT 1083	

RESULT 12

LOCUS AY088662 1352 bp mRNA linear PLN 14-APR-2003
DEFINITION Arabidopsis thaliana clone 8989 mRNA, complete sequence.
ACCESSION AY088662
VERSION AY088662.1 GI:21407436
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.I.
Full-length messenger RNA sequences greatly improve genome

annotation

Genome Biol. 3 (6), RESEARCH0029 (2002)

JOURNAL MEDLINE

PUBMED 22083475

REFERENCE 2 (bases 1 to 1352)

Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and

Feldmann,K.

Full-length cDNA from Arabidopsis thaliana

Unpublished to (1352)

Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and

Feldmann,K.

Direct Submission

Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,

Malibu, CA 90265, USA

This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to fLGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the Ws or Laer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genet carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.

FEATURES

source

1. .1352
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="8989"
80..1156
/codon_start=1
/product="cysteine proteinase AAP"

CDS

/protein_id="AA066984.1"
/db_xref="GI:21617934"
/translation="MSAKTILSSVVLVLAASAAANIGFDESPIRMVSDGLREVEE
SVSQILQSRHVLSPARFTHRYKKYQNVVEEMKLRFSIPKENLDLRSTNKKGLSYKL

GVNPFADLTWQEFORTKLGAAQNCATLKGSHKVTBAALPETKDWREBDGIVSPVQDG
CGCWTFTTGALERAYHQAQKGLSLSQQLVDCAGAFNNYGCNGGLPSQAPEYIK
SNGGLDTEKAYPTGKDETKFSAENYGVQVLSVNTLGAEBELKHAVALRPSVIA
FEVHSFRLYKSYVYTDHSCGSTPMDVNHAVLAVGVGVDPVFLIKNSWGDWGDK
GYFKWEMGXKNCGIATCASYPVVA"

ORIGIN

Query Match 45.6%; Score 488.4; DB 8; Length 1352;
Best Local Similarity 67.0%; Pred. No. 1.8e-123;
Matches 693; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY	32	TCCTGGTCTCCGCCCTCTCCGTCGCGCTCGCGGTTTCGAAGAAGACAAATCCAAATCCGGT	91
DB	120	TTCTCGTCCGCGCATCAGCAGCGGCTAATATCGGATTCGATGAGTCAAACCCGATCCGAA	179
QY	92	CGGTTACAAAGGCTGACTCGATTGAGCTGCCATCTCGGCGTCTTCGGCAGTTGCC	151
DB	180	TGGTCTCGATGGTCTCCGGAGGTAGAGAAATCTGTTTCCAGATCTTAGTCAATCTC	239
QY	152	GCCAGCGCTTCCACTTCGCACGGTTTCGCCCGCAGTACGGGAAGAGCTACGGATCGGAGG	211
DB	240	GTCAAGTTCTCTCTTCTGCTGCTTCACTCACCAGATATGTAAGAGTATCAGAACGTTG	299
QY	212	AGGAGATCAAGAGAGTTCCGATCTCGTGGAGAACTAGCGTTTATCCGTCCTCACTA	271
DB	300	AGGAGTGAAGTTCTCGATTCTCGATTTCAGGAAGATCTGATTGATGATGATCCACCA	359
QY	272	ATCGGAAGATCTGCTGATATACCTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGG	331
DB	360	ACAAGAAAGGCTTACTTACAAACTCGGTGTTAATCAATTTGCTGATTGATGATGCAAG	419
QY	332	AATTCGGACCAATCGCCTTGTGCGGCGCAGAACTGCTCGCGACTGGCATGGAACC	391
DB	420	AGTTTCAAAGGACCAAGCTTGGTGTCTCGAATCTGCTGCCACTTTAAAGGCGAGCC	479
QY	392	ACCGGTTTCTGATGCGGTCTTCTGTAAACAGGAGATTGGAGGAGCAAGGATAGTGA	451
DB	480	ACAAGTCAAGAGCAGCTCTTCTGAAACAAAGAGCTGGAGAGAGATGATCGTTA	539
QY	452	GCCTGTAAAGGACCAAGAGCTGTGGATCTTCTGAGACTTTCAGTACTACTGAGCAC	511
DB	540	GTCCGTCAAAGATCAGGAGGTTTGGATCTTGTGGACATTCAGCAACTGAGGCTC	599
QY	512	TAGAGCTGCATATACAGCTTAACCTGGAAGAGCAGCATCATTTCTGAACAGCAACTTG	571
DB	600	TTGAGGAGCTTACCATCAGGCAATTTGAAAGGAATATCTCTCTGAGCAAGCTTG	659
QY	572	TGGACTGTGCTCAGCAATCAATACTTTGGATGCAATGGAAGTTTGCTTCCCAAGCT	631
DB	660	TGGATTGTGCTGAGCTTTCAATACTATGTTGCAATGGTGGCTTCTCTTCTCAAGCT	719
QY	632	TTGAATACGTTAAGTACAAATGGAGGCTACGACAGCAAGACTTATCCATACCTTGGTG	691
DB	720	TTGAATACATAAATAACAGGTTGGCTGACACAGAGAAAGCTTATCTTATACCGTGA	779
QY	692	TCAATGTAATCTGCAACTTCAACGAGAGAAATTTGGTGTCAAGGTCAATGATTCGATAA	751
DB	780	AAGATGAACTTGCAATTTTACGTGAAACGTTGGTGTACAAAGTCTTCAACTCAGTCA	839
QY	752	ACATCACCTTGGGTGCTGAGATGAGTTGAAGCATGAGTGGCTTGGTGGCTTCCAGTTA	811
DB	840	ACATTACTTGGGTGCTGAAGATGAATGAAGCATGCGTTGGATTGGTACGCGCCAGTAA	899
QY	812	CGGTTGATTTGAGTTGTGAAAGTTTCAATCTGTACAGAAAGGTGTATACAGAGTG	871
DB	900	GCATAGATTTGAGTTTATACATCTGTTCCGGCTTTACAGAGTGGAGTTTACACTGATA	959
QY	872	ACACTGTGGAAGAGATCCAAATGATGTAACACGAGTCTTTCGGTGGTTATGGAG	931
DB	960	GTCACTGTGGAAGTACTCAATGATGTGAACACGCGGTTTGGCGGTTGGTTATGGAG	1019
QY	932	TCGAGGACGGATTCCTTATTTGGCTCATCAAGAACTCATCGGTACAAATTTGGGTGACA	991

DB	1020	TTGAAGACGGTGTACATATTGGCTTATTAGAACTCATGGGAGCGGATTTGGGGGACA	1079
QY	992	ATGGCTACTTTTAAGATGAACATCGGCAAGAACATGTGTGGTGTTCACATTCGGCATCTT	1051
DB	1080	AAGTTACTTCAAGATGGAGATGGGGAAGAACATGTGTGGTATTGCTATCATGTGCATCAT	1139
QY	1052	ATCCCATTTGGCT	1065
DB	1140	ACCCCGTTGGCT	1153

RESULT 13

AF083703 LOCUS 1355 bp mRNA linear PLN 29-OCT-2002
DEFINITION Arabidopsis thaliana clone sps232 unknown mRNA.
ACCESSION AF083703
VERSION AF083703.1 GI:24417303

KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

Subfamily: viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1355)
AUTHORS Stracke, R. and Palme, K.
TITLE Signal Peptide Selection derived cDNAs from Arabidopsis thaliana
leaves and guard cells

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1355)

AUTHORS Stracke, R. and Palme, K.

TITLE Direct Submission

JOURNAL Submitted (12-AUG-1998) Max-Planck-Laboratorium in der

Max-Planck-Gesellschaft, Carl-von-Linne-Weg 10, Koeln D-50829,

FEATURES

source

1. 1355

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="sps232"

/notes="signal peptide selection derived"

66..1142

/codon_start=1

/product="unknown"

/protein_id="AAN60262.1"

/db_xref="GI:24417304"

/translations="MSAKTILSSVWLVLVAASAAANIGPDSNPIRMVSDGLREVVE
SVQILGOSRHVLSFAFTHRYGKYQNVEMKLRFSI FKENLDLIRSTNKGILSYKL
GVNQFADLTWQEFORTKLGAAQNCATLKGSHKVTBAALPETKDWREBDGIVSPVQDG
CGCWTFTTGALERAYHQAQKGLSLSQQLVDCAGAFNNYGCNGGLPSQAPEYIK
SNGGLDTEKAYPTGKDETKFSAENYGVQVLSVNTLGAEBELKHAVALRPSVIA
FEVHSFRLYKSYVYTDHSCGSTPMDVNHAVLAVGVGVDPVFLIKNSWGDWGDK
GYFKWEMGXKNCGIATCASYPVVA"

CDS

ORIGIN

Query Match	45.6%;	Score 488.4;	DB 8;	Length 1355;
Best Local Similarity	67.0%;	Pred. No. 1.8e-123;	Indels 0;	Gaps 0;
Matches	693;	Conservative 0;	Mismatches 341;	
QY	32	TCCTGGTCTCCGCCCTCTCCGTCGCGCTCGCGGTTTCGAAGAAGACAAATCCAAATCCGGT	91	
DB	106	TTCTCGTCCGCGCATCAGCAGCGGCTAATATCGGATTCGATGAGTCAAACCCGATCCGAA	165	
QY	92	CGGTTACAAAGGCTGACTCGATTGAGCTGCCATCTCGGCGTCTTCGGCAGTTGCC	151	
DB	166	TGGTCTCGATGGTCTCCGGAGGTAGAGAAATCTGTTTCCAGATCTTAGTCAATCTC	225	
QY	152	GCCAGCGCTTCCACTTCGCACGGTTTCGCCCGCAGTACGGGAAGAGCTACGGATCGGAGG	211	
DB	226	GTCACTGTCTCTCTTCTGCTGCTTCACTCACCAGATATGTAAGAGTATCAGAACGTTG	285	
QY	212	AGGAGATCAAGAGAGTTCCGGATCTTCGTGGAGAACTCATCGGTATTCGGTCCACTA	271	

Db 286 AGGAGATGAAGCTTCGATTCCTGATTTTCAAGGAGAAATCTTGATTTGATCAGATCCACCA 345
 QY 272 ATCGAAGAGATCTGTGTATACCTAGGATCAACCAATTCGCGACCTGACTGGAGG 331
 Db 346 ACAAGAAGAGCTTATCTTCAAACTCGGTGTATCAATTTGCTGATTTGATGACATGGCAAG 405
 QY 332 AATTCCGGACCAATCGCTTGTGCGGCGCAGAACTGCTCGGCGACTGGCGATCGGAACCC 391
 Db 406 AGTTTCAAGGACCAAGCTTGTGCTCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCC 465
 QY 392 ACCGTTTGTGATGGGCTCTCTCTGTACAGAGGATTTGGAGGAGCAGGATAGTGA 451
 Db 466 ACAAGGTACAGAGAGCAGCTCTCTCTGAAACAAAGACCTGGAGAGAGATGTATCGTTA 525
 QY 452 GCCTGTAAAGGACCAAGGAGCTGTGATCTTCTGCGACTTTTCAGTACTTACTTGGAGCAC 511
 Db 526 GTCCGTTCAAGATCAGGAGGTTGTGATCTTCTGCTGGCAATTCACACAACTGGAGCTC 585
 QY 512 TAGAGCTGCTATACACAGCTAACTGGAAGAGCACATCATTTATCTGAACAGCAACTTG 571
 Db 586 TTGAGGAGCTTACCATCAGGCATTTGGAAAGGAATATCTCTCTGAGCAACAGCTTG 645
 QY 572 TGGACTGTGCTCAGCATCAATTAACCTTTGGATGCAATGAGGTTTGCCTTCCCAAGCCT 631
 Db 646 TGGATTGTGCTGAGCTTCAATAACTATGTTGCAATGTTGCTTCCCTTCTCAAGCCT 705
 QY 632 TTGAATACGTTAAGTACAATGAGGACATCGACAGACAGACTTATCCATACCTTTGGTG 691
 Db 706 TTGAATACATAAATCAACGCTGCTCGACACAGAGAACTTATCTTATACCGGTA 765
 QY 692 TCAATGATCTGCAACTCAAGCAGGAGAACTGTTGGTGTCAAGTCAATGATTCGATTA 751
 Db 766 AAGATGAACCTGCAATTTTACGTGAAACGTTGGTGTACAGTCTCAACTCAGTCA 825
 QY 752 ACATCACCTCGGTGCTGAGGATGAGTTGAAGCATGCAGTGGCTTGGTGGCTCCAGTTA 811
 Db 826 ACATTACTCTGGTGTGAGTGAAGTGAAGTGAAGTGGTGGTGGTGGTGGTGGTGG 885
 QY 812 GCGTTGATTTGAGTTGGAAGGTTTCAATCTGTACAGAAAGGTGTATACAGAGTG 871
 Db 886 GCATAGCATTTGAGTTTATACATCTCGTTCGGCTTTTACAAGAGTGGAGTTTACACTGA 945
 QY 872 ACACCTGTGAAGAGATCCATGATGTGAACACGAGTCTTTCGCTCGGTTATGGAG 931
 Db 946 GTCACTGTGAAGTACTCCATGATGTGAACACGCGGTTTGGCGGTTGTTATGGAG 1005
 QY 932 TCGAGGACGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGTGACA 991
 Db 1006 TTGAAGACGGTGTACCATATTGGCTTATTAAAGTCAATGGGAGCGGATTGGGGCGACA 1065
 QY 992 ATGGCTACTTTAGATGGAATCGGCAAGACATGTGGTGTGCAACTTGGCATCTT 1051
 Db 1066 AAGTTTACTTCAAGATGGAGTGGGAAGAACATGTGTGTTATTGCTATGTGCAATCAT 1125
 QY 1052 ATCCCATTTGGCT 1065
 Db 1126 ACCCGTTTGGCT 1139
 RESULT 14
 LOCUS BT000676 1331 bp mRNA linear PLN 01-OCT-2002
 DEFINITION Arabidopsis thaliana clone RAFLO8-12-G17 (R11172) putative cysteine
 proteinase AALP (At5g60360) mRNA, complete cds.
 ACCESSION BT000676
 VERSION BT000676.1 GI:23397073
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1 (bases 1 to 1331)

AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
 Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C.,
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
 Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
 Ecker,J.R. and Theologis,A.
 TITLE Arabidopsis Full Length cDNA Clones
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1331)
 AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
 Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
 Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R.,
 Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J.,
 Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
 Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 TITLE Direct Submission
 JOURNAL Submitted (01-OCT-2002) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN
 Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
 Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M.,
 Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L.,
 Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C.,
 Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
 Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
 Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to
 this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC)
 contributed equally to this work as PIs.

Annotation based on January 2002 version of the Arabidopsis genome
 submitted to Genbank.

FEATURES
 source

1..1331
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="RAFL08-12-G17 (R11172)"
 /ecotype="Columbia"
 /note="This clone is in a modified pBluescript vector
 (FLC-1) as a BamHI/XhoI insert."

gene

1..1331

5'UTR

/gene="At5g60360"

misc_difference

2..4

/gene="At5g60360"
 /note="compared to genomic sequence"
 /replace="gct"
 69..1145
 /gene="At5g60360"
 /codon_start=1
 /evidence=experimental
 /product="putative cysteine proteinase AALP"
 /protein_id="AA031822.1"
 /db_xref="GI:23397074"
 /translation="MSAKTILSSVVLVVAASAAANIGFDESPIRMVSDGLREVVEE
 SVSQTIGOSRHLVSPFTHRYKQYQVEMKLFSPFENLIDIRSTNNKGLSYKL
 GYNQPADLTNQEFORTKLGAQNGCSATLKGSHKVTAEALPETKQWREGIVSPVQDQ
 GCGSWTFTTCAEAHQAFQKGISLSEQLVDCAGAFNNYGGNGGLPSQAFYIK
 SNGGLDTKAPYPTGKSTCFPSAENVGVQVLSNVITLGADELKHAUGLVRPVSIA
 FEVTHSPFLYSGVYTDHSGSTPMNVHVAVLAVGYGVEDGVFVWLLIKNSWGADWGDK
 GYFKMEMGKNMGCIATCASYPVA"
 1146..1331

CDS

/gene="At5g60360"

3'UTR

1146..1331

ORIGIN /gene="At5g60360"

Query Match 45.5%; Score 486.8; DB 8; Length 1331;
 Best Local Similarity 66.9%; Pred. No. 5e-123;
 Matches 692; Conservative 0; Mismatches 342; Indels 0; Gaps 0;

```

QY 32 TCCTGGTCTCCGCCCTCTCCGTCGCGTCGCCGTTTCGGAAGAAGACAATCCAAATCCGGT 91
DB 109 TTCTCGTCGCGCATCAGCAGCGGCTAATATCGAATTCGATGASTCAAAACCCGATCCGAA 168

QY 92 CCGTTACACAAAGCCCTGACTCGATTGAGCCTGCCATCTCGCGCTCTCTGGCAGTTGCC 151
DB 169 TGGTCTCCGATGGTCTCCGGAGGTAGAAGATCTGTTCCAGATCTTAGGTCAATCTC 228

QY 152 GCCACGCTTCCACTTCGCACGGTTCCCGCGCAGGTACGGGAAGAGCTACGATCCGAGG 211
DB 229 GTCAGGTTCTCTCTTTCGCTCGTTCACTCACCGATATGTTAAAGATATCAGACGTGG 288

QY 212 AGGAGATCAAGAAGAGGTTCCGGATCTTCCTGGAGATCTAGCGTTTATCCGCTCCACTA 271
DB 289 AGGAGATGAAGCTTCGATTTCTCGAATTTCAAGGAGAACTTTGATTTGATCAGATCCACC 348

QY 272 ATCGGAAGGATCTGCTGTTATACCTTAGGAATCAACCAATTCGCGCAGCTCACCTGGGAG 331
DB 349 ACAGAAGGCTTATCTTACAACTCGTGTAACTGATCAATTTGCTGATTTGACATGGCAAG 408

QY 332 AATTCGGACCAATTCGCTGGTGGCGCAGAACTGCTCGGCGACTGCGCATGGAAACC 391
DB 409 AGTTTCAAGGACCAACAGCTTGGTCTGCTCAGAACTGCTCTGCCACTTTTAAAGGGCAGCC 468

QY 392 ACCGGTTTTCGATGGGTGCTTCTCTGTAACGAGGGATTCGAGGGAGCAAGGATAGTGA 451
DB 469 ACAAGTTCACAGAGCAGCTTCTCTCAACAAAGACTCGAGAGAAGATGGTATCGTTA 528

QY 452 GCCCTGTAAGGACCAAGGAAGCTGTGGATCTTCTGGAATCTTTCAGTACTCTGGAGCAC 511
DB 529 GTCCGGTCAAGATCAGGAGGTTGTGGATCTTCTCTGGACATTCAGCACAACTGGAGCTC 588

QY 512 TAGAGGCTGCATATACACAGCTACTGGAAGAGCACATCATTTATCTGAACAGCAACTGG 571
DB 589 TTGAGGAGCTTTACATCAGGCAATTTGGAAAGAAATTTCTCTCTGAGCAACAGCTTG 648

QY 572 TGGACTGTGCTCAGCAATTCATAAATTTTGGATCAATTCGAGGTTTGCCTTCCCAAGCCT 631
DB 649 TGGATTGTCTGGAGCTTTCATTAACATATGGTTGCAATGGTGGCCCTTCTCTCAAGCCT 708

QY 632 TTGAATAGCTTTAAGTACAAATGGAGGCATCGACACAGAACAGCTTATTCATCTCTGGTG 691
DB 709 TTGAATACATCAAAATCCAAACGGTGGCTTCGACACAGAGAAAGCTTATCTCTTATACCG 768

QY 692 TCAATGGTATCTGCAACTTCAGCAGGAGAAATGTGGTGTCAAGGTCAITGATTTCGATAA 751
DB 769 AAGATGAACCTGCAATTTTCAGCTGAAACGTTGGTGTACAAGTCTCAACTCAGTCA 828

QY 752 ACATCACCTCGGTGCTGAGGATGAGTTGAAGCATGCACTGGGCTTGGTGGCTCCAGTTA 811
DB 829 ACATTAATCTCGGTGCTGAAGTGAAGTGAAGTGGGTTGGATGGTACGGCCAGTAA 888

QY 812 CGGTGCAATTTGAGGTTCTGAAAGGTTTCAATCTGTACAGAAAGGTGTATACAGCAGTG 871
DB 889 GCATAGCAATTTGAGGTTATACACTCGTTCGGCTTACAGAGTGGAGTTTACACTGNTA 948

QY 872 ACACCTGTGGAAGAGATCCAAATGGATGTGAACACGACAGTTCTTGGCGCTCGGTTATGAG 931
DB 949 GTCACGTGGAAGTACTCCAAATGGATGTGAACCAACGCGGTTTTCGGCGCTTGGTTATGGAG 1008

QY 932 TCGAGGACGGGATTCCTTATTTGGCTTCATCAAGAACTCATGGGTACAAATTTGGGTTGACA 991
DB 1009 TTGAAGACGGGTGACCAATTTGGCTTATTAAAGACTCATGGGAGCGGATTTGGGCGACA 1068

QY 992 ATGGCTACTTTAAGATGAACTCGGCAAGAACATGTGTGGTGTGTTTCCAACTTTCGCACTCT 1051

```

5'UTR 1. .39 /gene="At5g60360"
misc_difference 1. .2 /gene="At5g60360"
/note="compared to genomic sequence"
/replace="tg"
misc_difference 4 /gene="At5g60360"
/note="compared to genomic sequence"
/replace="g"
40. .1116 /gene="At5g60360"
/note="contains a papain family cysteine protease domain
PF00112"
/codon_start=1
/evidence=experimental
/product="putative cysteine proteinase AAPL"
/protein_id="AAK25983.1"
/db_xref="GI:13430722"
/translation="MSAKTILSSVLVLAASAAANIGPDESPIRMVSDGLREVEE
SVQILGQSRHVLVSFAFTHRYGKYQNVSEMKLRSIFKENLDLRSINTKGLSKL
GVNQPADLTWQEFQRTKLGAQNCSTALKGSHVTEAALPETKDWREDGIVSVQDQ
GCSCWTFSTTGALEAHQAFQKISLSEQQLVDCAGAFNFCNGGLPSPQAFYIK
SNGGLDTEKAVPTVTKDCTKFSANVGVOLANSVNTLLGADELKHAVGLVRPVSIA
FEVTHSPRLYKSGVYTDHSGSTPMDVNHAVLAVGVEDGVFYLKNSWGADWDK
GYFKMENGKMGIAATCASYPVA"
1117. .1370
/gene="At5g60360"
1354
misc_feature /gene="At5g60360"
/note="artifact within poly A tail"

ORIGIN

Query Match 45.5%; Score 486.8; DB 8; Length 1370;
Best Local Similarity 66.9%; Pred. No. 5e-123;
Matches 692; Conservative 0; Mismatches 342; Indels 0; Gaps 0;
QY 32 TCCTGGTCTCGCCCTCTCGTCCGCTCGCGGTTTCGAAGAAGACAAATCCATCCGGT 91
Db 80 TTCTGTCGCCGATCAGCAGCGGCTAATTCGGAATTCGATGAGTCAAAACCGGATCGAA 139
QY 92 CCCTTACACAAAGCCGTGACTCGATTGAGCTGCCATCCCTCGCGGCTCTTGGCAGTTGCC 151
Db 140 TGGTCTCGATGGTCTCCGGGAGGTAGAGAAATCTGTTCCAGATCTTAGGTCAATCTC 199
QY 152 GCCAGCCTTCCATTCGCACGGTTCGCCCGCAGGTACGGGAGAGCTACGGATCGGAGG 211
Db 200 GTACAGTTCTCTCTCTCGCTCGCTTCACTACCGGATATGGTAAAGATATCAGAACGTGG 259
QY 212 AGGAGATCAAGAAGAGGTTCGGGATCTTCGTGGAGAACTTAGCGTTTATCGGTCCACTA 271
Db 260 AGGAGATGAAGTTCGATTCGATTTTCAAGGAGAACTTTGATTCAGATCCACCA 319
QY 272 ATCGAAGGATCTGTGTAATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGG 331
Db 320 ACAAGAAAGGCTTATCTTACAAATTCGGGTGTTAATCAATTTGCTGATTTGACATGGCAAG 379
QY 332 AATTCGGACCAATCGCTTGGTGGGCGCAGAACTGCTCGCGCACTCGCGCATCGAAACC 391
Db 380 AGTTTCAAGAGCAAGCTTGGTGTCTCAGACTGCTCTGCCACTTAAAGGGCAGCC 439
QY 392 ACCGGTTGTTCGATGGCGTGTCTCTGTAAAGAGGATGGAGGAGCAAGGATAGTGA 451
Db 440 ACAAGGTACAGAAAGCAGCTTTCCTGAAACAAAGAGCTGGAGAGAAAGATGGTATCGTTA 499
QY 452 GCCCTGTAAGGACCAAGGAGCTGTGGATCTTGTGGACTTTCAGTACTCTGAGCAC 511
Db 500 GTCCGGTCAAAGATCAGGAGGTTGTGGATCTTGTGGACATTCAGCACTCGAGCTC 559
QY 512 TAGAGGCTGCATATACAGCTTAATCGAAAGAGCACATCAATATCTGAAACAGCAACTTG 571
Db 560 TTGAGGCGAGCTTACCATTGAGGCAATTTGAAAAGGAATTTCTCTCTCTGAGCAACAGCTTG 619

QY 572 TGGACTGTGCTCAGCATTCAATTAACCTTTGATGATGGAGGTTTGCTTCCCAAGCCT 631
Db 620 TGGATTGTGCTGGAGCTTTCAATAAATATGTTGCAATGGTGGGCTTCCCTTCTCAAGCCT 679
QY 632 TTGAATACGTTAAGTACAAATCGAGGCATCGACACAGAAACAGACTTATCCATACCTTGGTG 691
Db 680 TTGAATACATCAATCCAAACGGTGGCTCGACACAGAGAAAGCTTATCCTTATACCGGTA 739
QY 692 TCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGTCAATGATTCGATAA 751
Db 740 AAGATGAAACCTGCAAAATTTTTCAGCTGAAACAGTTTGGTGTCAAGTCCCTCAACTCAGTCA 799
QY 752 ACATCACCTCTGGTCTCGAGGATGATTGAAGCATGCAAGTGGGCTTGGTGGTCCAGTTA 811
Db 800 ACATTACTCTGGTCTGAGATGACCTGAAGCATCGCGTTGGATTGGTACGGCCAGTAA 859
QY 812 GCGTTGCATTTGAGGTTGTGAAAGGTTTCAATCTGTACAAGAAAGGTGTATACAGCATG 871
Db 860 GCATAGCAATTTGAGGTTATACACTCGTTCGGCTTTACAAGAGTGGAGTTTACACTGATA 919
QY 872 ACACCTGTGGAAGATCCAAATGGATGTGAACCAAGCAGTTCTTCCCGTTCGGTTATGGAG 931
Db 920 GTCACTGTGGAAGTACTCCAAATGGATGTGAACCAAGCAGTTTGGCCGTTGGTTATGGAG 979
QY 932 TCGAGGACGGGATTCCTTATTTGGCTCATCAAGAACTCATGGGTTACAAATTTGGGTTGACA 991
Db 980 TTGAAGACGGTGTACCATATTGGCTTATTAGAACTCATGGGAGCGGATTTGGGCGACA 1039
QY 992 ATGGCTACTTTAAGATGGAACCTCGGCAAGACATGTGGTGTTCGAACCTTGGCATCTT 1051
Db 1040 AAGGTTACTTCAAGATGGAGATGGGGAAGAACATGTGGTATTGCTATCATGTGCATCAT 1099
QY 1052 ATCCCAATTTGGCT 1065
Db 1100 ACCCGTTTGGCT 1113

Search completed: March 29, 2004, 20:28:32
Job time : 2968 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 17:13:57 ; Search time 373 Seconds
(without alignments)

12197.909 Million cell updates/sec

Title: US-10-087-714-1

Perfect score: 1071

Sequence: 1 atggcagctaagctctctt.....atccattgtgctgtgtag 1071

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002s.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1071	100.0	1071	9	Adc61376 DNA encod
2	569.6	53.2	1442	2	Rat35111 Enhanced
3	541.4	50.6	1407	4	Aaf44802 Cysteine
4	541.4	50.6	1407	9	Adb94743 Programme
5	521.6	48.7	1467	2	Rav82457 Triticum
6	488.4	45.6	1358	3	Rac37877 Arabidops
7	486.8	45.5	1370	3	Rac48945 Arabidops
8	478.2	44.6	1363	3	Rac40628 Arabidops
9	475.8	44.4	1388	2	Rat35100 Up-regula
10	467.6	43.7	1077	3	Rac51435 Arabidops
11	467.6	43.7	1282	6	Abn98252 Arabidops
12	429.8	40.1	1072	3	Rac44535 Zea mays
13	429.4	40.1	2149	3	Rac44178 Arabidops
14	407.4	38.0	1652	9	Rdb94877 Programme
15	351.8	32.8	888	9	Rdb94874 Programme
16	349.8	32.7	675	8	ACL19836 DNA clone
17	348.8	32.6	618	8	ACL19838 DNA clone
18	347.8	32.6	662	8	ACL19837 DNA clone
19	327.6	30.6	590	8	ACL19847 DNA clone
20	327.2	30.6	630	8	ACL19848 DNA clone
21	323.4	30.2	565	8	ACL19840 DNA clone
22	321.6	30.0	644	8	ACL19841 DNA clone
23	305.8	28.6	604	8	ACL19843 DNA clone

C 24	304	28.4	536	8	ACL19835	AC119835 DNA clone
C 25	298.4	27.9	588	7	ABX56737	ABX56737 Arabidops
C 26	297.8	27.8	595	8	ACL19839	AC119839 DNA clone
C 27	293	27.4	578	8	ACL19845	AC119845 DNA clone
C 28	285.4	26.6	550	8	ACL19846	AC119846 DNA clone
C 29	283.4	26.5	545	8	ACL19849	AC119849 DNA clone
C 30	280.2	26.2	677	4	AAF44806	Aaf44806 Cysteine
C 31	280.2	26.2	677	9	ADB94747	Adb94747 Programme
C 32	278.2	26.0	499	8	ACL19831	AC119831 DNA clone
C 33	269.8	25.2	1362	7	ABT41873	Abt41873 Toxicity
C 34	269.8	25.2	1362	9	ADB52643	Abd52643 Primary r
C 35	257.6	24.1	714	8	ACL19871	AC119871 DNA clone
C 36	253.2	23.6	530	8	ACL19827	AC119827 DNA clone
C 37	249.4	23.3	609	8	ACL19854	AC119854 DNA clone
C 38	249.4	23.3	710	8	ACL19865	AC119865 DNA clone
C 39	248.6	23.2	718	8	ACL19870	AC119870 DNA clone
C 40	247.2	23.1	723	8	ACL19866	AC119866 DNA clone
C 41	247.2	23.1	1399	6	ABK84545	Abk84545 Human cDN
C 42	244.4	22.8	479	8	ACL19842	AC119842 DNA clone
C 43	244	22.8	1433	9	ADD18512	Add18512 Human pro
C 44	244	22.8	1475	4	AAI93479	Aai93479 Human pol
C 45	244	22.8	1475	4	AAI93494	Aai93494 Human pol

ALIGNMENTS

RESULT 1

ADC61376

ID ADC61376 standard; DNA; 1071 BP.

AC ADC61376;

DT 18-DEC-2003 (first entry)

XX DNA encoding Vanilla planifolia 4-hydroxybenzaldehyde synthase (4HBS).

KW Vanillin production; plant; vanillin biosynthesis;

XW 4-hydroxybenzaldehyde synthase; 4HBS; gene; ds.

OS Vanilla planifolia.

XX Key Location/Qualifiers

FT CDS 1..1071

FT /*tag= a

FT /product= "4HBS"

FT /transl_except= (pos:481..489,aa:Ser-Trp)

FT /transl_except= (pos:538..546,aa:Gly-Ser)

FT /transl_except= (pos:547..555,aa:Thr-Leu)

FT /transl_except= (pos:604..612,aa:Cys-Gly)

FT /transl_except= (pos:694..696,aa:Met)

XX US2003070188-A1.

XX 10-APR-2003.

XX 28-FEB-2002; 2002US-00087714.

XX 15-JUL-1997; 97US-0052604P.

XX 15-JUL-1996; 98WO-US014895.

XX 25-MAY-2000; 2000US-00462576.

XX 28-FEB-2001; 2001US-0272415P.

XX (HVK/) HAVKIN-FRENKEL D.

XX (PODS/) PODSTOLSKI A.

XX (DIXO/) DIXON R. A.

XX Havkin-Frenkel D, Podstolski A, Dixon RA;

XX WPI; 2003-743794/70.

XX P-PsDB; ADC61377.

XX Improving vanillin production in cultured Vanilla planifolia by

21 CTTCTACTCTTCTCTGCTCTCGCCCTCTCGTGGCGCTCGCGGTTTCGAGAGACAA 80
Db
87 CGTCTCTCTCTCTGCTGCTGCTCTCGCGCGCTCGAGCTTCGAGAGTCCAA 146
Qy
81 TCCATATCGGTCGCTTACACA---AAGGCTTACCTGAGCTGAGCTGCTCGCGGT 137
Db
147 CCCCATCGGCTCTTCCCGAGCGGCGCTCGGACCTCGAGTCTCTCAATCGTCCAGAT 206
Qy
138 CTTGGCAGTTTCGCGCCAGCGCTTCCATTTGCGACGTTCTCCCGCGAGTACGGGAGAG 197
Db
207 CGTGGCGCGCACCGCCACCGCTTCTCTTGGCGCGCTTCGCCAACAGGTATGGGAAGAG 266
Qy
198 CTACGATCGGAGGAGATCAAGAGAGTTTCGGATCTTCGTGGAGATCTTAGCGTT 257
Db
267 GTACGAGACCGGAGGAGATCAAGCTGGGTTTCAGATCTTCAGGAGATCTCAAGTT 326
Qy
258 TATCGGCTTCCATATCGGAGGATCTGTGATATACCTTAGGAATCAACCAATTCGCGGA 317
Db
327 GATCGATCCACCAAGAGGCTTGCCTTACACCTCGGTGTCATTAAGTTTCTGA 386
Qy
318 CCGTACCTGGAGGAAATTCGAGACCAATCGCTTGTGGGCGAGAACTCTCGGCGAC 377
Db
387 TTGGAGCTGGAGGATTCAGAGGCACAGACTCGGAGCTGCTCAAACTGCTCTGCCAC 446
Qy
378 TCGCATGCAAAACCGGCTTGTGATGGCGTCTCTGTAAACGAGGATTCGAGGGA 437
Db
447 CACCAAGGCAACCAAGCTCACGACGAAGCTCTTCCGAGATGAAGACTGGAGAGA 506
Qy
438 GCAAGGATGAGAGCCCTGTAAGGACCAAGAGAGCTGTGATCTTCTGAGACTTTCAG 497
Db
507 AAGGGCATTTGAAGCCCAATTAAGATCAGGGGCACTGTGATCTTCTGAGCTTTCAG 566
Qy
498 TACTACTGAGCACTAGGCTGCAATATACAGCTTAACGAAAGAGACATCATATATC 557
Db
567 TACCTTGGAGCTTTGAGCTGCTTATCACCAGCAATTCGGGAACAAATCTCTCTGC 626
Qy
558 TGAACAGCACTTGTGAGCTGTGCTCAGCATTCATTAACCTTGAATGCAATGAGGTTT 617
Db
627 TGAGCAGCAGCTCGTGAGCTGTCTGGGCTTTCACCACTTTGATGTAGTGTGAGT 686
Qy
618 GCCTTCCCAAGCTTTGATAGTTAAGTAACTGAGGAGCTGACAGACAGACTTA 677
Db
687 GCCATCCCAAGCTTTGATAGTCAAGTAACTGAGGAGCTGAGTACGAGAGAGATA 746
Qy
678 TCCATACCTTGTGTCAATGCTATCTGCACTTCAAGCAGGAGATTTGTGTCAAGT 737
Db
747 TCCTTATACGCACTGATGTTGCTGCAATTTCTCGGCTGATATGTTGTGTCCAAAT 806
Qy
738 CATGATTCGATAAATCAATCCCTGGGCTGTGAGGATGATTTGAAGCATGCAATGGCTT 797
Db
807 GCTCGACTCTGTAAACATCACCTTGGGTGTGAGGATGAACATAAGCATGCAATGCTT 866
Qy
798 GGTGCGTCCAGTTAGGTTGCTTTCAGGTTGTGAAAGTTTCAATCTGTACAAAGAG 857
Db
867 CGTCCGCGCAGTGTGCTTTCAGGCTGCTGAAAGCTTCAGATTCAGTTCGGG 926
Qy
858 TGTATACAGCAGTGACACTGTGGAAGAGATCAATGATGTAACACGAGCTTCTTGC 917
Db
927 TGTCTACAGCAGGATACATGCGGTAGCACTTCCATGATGTGAACATGCTTCTCGC 986
Qy
918 CGTGGTTATGAGTTCGAGGACGGATTCCTTATTCGCTCATCAAGAACTCATGCGGTAC 977
Db
987 TGTGTTATGAGTTGAGATGTTTTCGTTCTGGCTCATCAAGAAATTCCTGGGAGC 1046
Qy
978 AAATTTGGGTGACAATGGCTACTTTTAAAGTGGAACTCGGCAAGAAATGTGTGTTGC 1037
Db
1047 AGACTGGGTGACCAAGGATCTCAAGATGGAGATGGGAAGAAAGATGTGTGAGTCCG 1106
Qy
1038 AACTTGGCATCTTATCCCATTTGGC 1064
Db
1107 TACTTGTGCATCATACCTTGTGGC 1133

RESULT 4
ADB94743
ID ADB94743 standard; cDNA; 1407 BP.
XX
AC ADB94743;
XX
DT 04-DEC-2003 (first entry)
XX
DE Programmed cell death pathway cysteine protease cDNA #16.
XX
KW Programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
KW Dnase; lls; lsd; nucellin-like aspartic protease; annexin; prohibitin;
KW fen-like protein; rac2; retinoblastoma-related protein; SINA; TFIID;
KW TGT; xylogenic RNase; pur-alpha; cyteine protease; RPP5-like protein;
KW GP 91 NADPH oxidase subunit; NRP-like protein; BAG-1;
KW defender against cell death; lethal leaf spot; lesion stimulating death;
KW seven in absentia; transcription initiation factor;
KW testis enhanced gene transcript; gene; ss.
XX
OS Eucalyptus grandis.
XX
XX US2003082724-A1.
XX
XX 01-MAY-2003.
XX
XX 14-AUG-2002; 2002US-00219220.
XX
XX 04-JUN-1999; 99US-00325932.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Flinn B, Lasham A;
XX
XX WPI; 2003-786916/74.
XX
XX New isolated polynucleotide useful for modulating programmed cell death,
XX altering the development cycle of plant cells, and subsequently modifying
XX plant development.
XX
XX Claim 1; Page 71-72; 214pp; English.
XX
XX The invention describes an isolated polynucleotide (I) comprising a
XX nucleotide sequence that is one of 145 fully defined sequences of 221-
XX 3415 base pairs (bp), given in the specification, and/or its complements,
XX reverse complements, reverse sequences, or sequences having 75, 90 or 95
XX % sequence identity to or that hybridise under stringent hybridisation
XX conditions to one of the 145 sequences. The methods and compositions of
XX the present invention to do with (I) are useful for modulating programmed
XX cell death and thereby altering the development cycle of plant cells, and
XX altering plant development. This sequence encodes a protein associated
XX with the programmed cell death pathway.
XX
SQ Sequence 1407 BP; 349 A; 345 C; 377 G; 336 T; 0 U; 0 Other;
Query Match 50.6%; Score 541.4; DB 9; Length 1407;
Best Local Similarity 70.5%; Pred. No. 6.6e-153;
Matches 738; Conservative 0; Mismatches 306; Indels 3; Gaps 1;
Qy 21 CTTCTACTCTTCTGCTCTCGCCCTCTCGTGGCGCTCGCGGTTTCGAGAGACAA 80
Db 87 CGTCTCTCTCTGCTGCTGCTCTCGCGCGCTCGAGCTTCGAGAGTCCAA 146
Qy 81 TCCATATCGGTCGCTTACACA---AAGGCTTACCTGAGCTGAGCTGCTCGCGGT 137
Db 147 CCCCATCGGCTCTTCCCGAGCGGCGCTCGGACCTCGAGTCTCTCAATCGTCCAGAT 206
Qy 138 CTTGGCAGTTTCGCGCCAGCGCTTCCATTTGCGACGTTCTCCCGCGAGTACGGGAGAG 197
Db 207 CGTGGCGCGCACCGCCACCGCTTCTCTTGGCGCGCTTCGCCAACAGGTATGGGAAGAG 266
Qy 198 CTACGATCGGAGGAGATCAAGAGAGTTTCGGATCTTCGTGGAGATCTTAGCGTT 257
Db 267 GTACGAGACCGGAGGAGATCAAGCTGGGTTTCAGATCTTCAGGAGATCTCAAGTT 326

Db 700 GGCATTTGAGTACATCAATAATGAGGAGATCGATACCGAGGAGTCTTACCCTTACAA 759
QY 687 TGGTGTCAATGATATCTGCAACTCAAGCAGAGAAATGTTGGTTCGAAGGTCAATGATTC 746
Db 760 CGGTGTCAATGGCGTCTGCAATACAAAGCTGAAATCCGTAGTTCAGGTTTGGACTC 819
QY 747 GATAAATCACCCTGGGTGCTGAGGATGAGTTGAAGCATGCAAGTGGCTGGTGGCTGCC 806
Db 820 TGTTAACATCACTGAATGCTGAGACGAACCTGAAACGCCGCTGGTGGTTCGCCCC 879
QY 807 AGTTAGCGTTGCAATTTGAGTTGTTGAAAGTTTCAATCTGTACAAAGAGTGTATACAG 866
Db 880 AGTGAAGTGTGCTTGGTGGTATCAACCGTTTCAGGCGATACAAAGCGGAGTTTACAC 939
QY 867 CAGTGACACCTGGAAGAGATCCATCGATGATGCAACCGAGTCTTGGCGTGGTTA 926
Db 940 TAGCGACCATGCGGTACTACACCTGATGACGTGAACCAACCGCTTGGCGGTGGGTA 999
QY 927 TGGAGTCGAGGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGG 986
Db 1000 CGGGTGGAGAACGGCGTCCCGTATTGGCTCATCAAGAACTCGTGGGGCGGCGACTGGGG 1059
QY 987 TGACAATGCTACTTTAATGATGGAATCGGCAAGACATGTTGCTGTGCACTTGGCGC 1046
Db 1060 TGACAATGCTTACTTCAAGATGGAATGGGCAAGAACATGTGTGCTGTGCTGTGCTGTGC 1119
QY 1047 ATCTTATCCCATTTGGCTG 1066
Db 1120 GTTCCTACCCCATCGTAGCGG 1139

RESULT 6
AAC37877
ID AAC37877 standard; DNA; 1358 BP.
AC AAC37877;
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 18989.
DE Arabidopsis thaliana
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 01-APR-1999; 99US-0126785P.
PR 06-APR-1999; 99US-0127462P.
PR 08-APR-1999; 99US-0128234P.
PR 16-APR-1999; 99US-0128714P.
PR 19-APR-1999; 99US-0129845P.
PR 23-APR-1999; 99US-0130077P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130691P.
PR 30-APR-1999; 99US-0131149P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132488P.
PR 11-MAY-1999; 99US-0132489P.
PR 14-MAY-1999; 99US-0132490P.
PR 14-MAY-1999; 99US-0132491P.
PR 14-MAY-1999; 99US-0132492P.
PR 18-MAY-1999; 99US-0132493P.
PR 18-MAY-1999; 99US-0132494P.
PR 20-MAY-1999; 99US-0132495P.
PR 21-MAY-1999; 99US-0132496P.
PR 24-MAY-1999; 99US-0132497P.
PR 25-MAY-1999; 99US-0132498P.
PR 27-MAY-1999; 99US-0132499P.
PR 28-MAY-1999; 99US-0132500P.
PR 01-JUN-1999; 99US-0132501P.
PR 03-JUN-1999; 99US-0132502P.
PR 04-JUN-1999; 99US-0132503P.
PR 07-JUN-1999; 99US-0132504P.
PR 08-JUN-1999; 99US-0132505P.
PR 10-JUN-1999; 99US-0132506P.
PR 10-JUN-1999; 99US-0132507P.
PR 14-JUN-1999; 99US-0132508P.
PR 16-JUN-1999; 99US-0132509P.
PR 16-JUN-1999; 99US-0132510P.
PR 17-JUN-1999; 99US-0132511P.
PR 18-JUN-1999; 99US-0132512P.
PR 18-JUN-1999; 99US-0132513P.
PR 18-JUN-1999; 99US-0132514P.
PR 18-JUN-1999; 99US-0132515P.
PR 18-JUN-1999; 99US-0132516P.
PR 18-JUN-1999; 99US-0132517P.
PR 18-JUN-1999; 99US-0132518P.
PR 21-JUN-1999; 99US-0132519P.
PR 23-JUN-1999; 99US-0132520P.
PR 23-JUN-1999; 99US-0132521P.
PR 24-JUN-1999; 99US-0132522P.
PR 28-JUN-1999; 99US-0132523P.
PR 29-JUN-1999; 99US-0132524P.
PR 30-JUN-1999; 99US-0132525P.
PR 01-JUL-1999; 99US-0132526P.
PR 01-JUL-1999; 99US-0132527P.
PR 02-JUL-1999; 99US-0132528P.
PR 06-JUL-1999; 99US-0132529P.
PR 08-JUL-1999; 99US-0132530P.
PR 09-JUL-1999; 99US-0132531P.
PR 12-JUL-1999; 99US-0132532P.
PR 13-JUL-1999; 99US-0132533P.
PR 14-JUL-1999; 99US-0132534P.
PR 15-JUL-1999; 99US-0132535P.
PR 16-JUL-1999; 99US-0132536P.
PR 16-JUL-1999; 99US-0132537P.
PR 19-JUL-1999; 99US-0132538P.
PR 19-JUL-1999; 99US-0132539P.
PR 19-JUL-1999; 99US-0132540P.
PR 19-JUL-1999; 99US-0132541P.
PR 20-JUL-1999; 99US-0132542P.
PR 20-JUL-1999; 99US-0132543P.
PR 20-JUL-1999; 99US-0132544P.
PR 21-JUL-1999; 99US-0132545P.
PR 21-JUL-1999; 99US-0132546P.
PR 21-JUL-1999; 99US-0132547P.
PR 22-JUL-1999; 99US-0132548P.
PR 22-JUL-1999; 99US-0132549P.
PR 22-JUL-1999; 99US-0132550P.

```
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 27-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145318P.
PR 27-JUL-1999; 99US-0145911P.
PR 27-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.

PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 45.6%; Score 488.4; DB 3; Length 1358;
Best Local Similarity 67.0%; Pred. No. 7.2e-137;
Matches 693; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY 32 TCCTGGTCTCCGGCTCTCCGTCCGGCTCCGGCTTCGAAAGAGACAAATCCATCCGGT 91
DB 121 TTCTCGTCGGCGCATCAGCAGCGGCTATATCGGATTCGATGAGTCAAAACCGGATCCGAA 180
QY 92 CGGTTACACAAAGGCTGACTCGATTGAGCCTCCGCTCCGGCTCTCCGGCTTCGCCAGTTC 151
DB 181 TGGTCTCCGATGGTCTCCGGAGGTTAGAAATCTGTTTCCAGATCTTAGTCAATCTC 240
QY 152 GCCAGCCTTCACATTCGCACGGTTCCGCCGAGTACGGAGAGAGTACGGATCGGAGG 211
DB 241 GTCACGTTCTCTCTTTCGTCGTTCACTCACCGATATGTTAAAGATATCAGAACGTTGG 300
QY 212 AGGAGATCAAGAGAGTTCCGGATCTTCGGTGGAGATCTAGCGTTTATCCGGTCCACTA 271
DB 301 AGGAGATGAAGCTTCGATTCGATTTTCAAGGAGAACTTTGATTTGATCAATCCACA 360
QY 272 ATCGGAAGGATCTGCTGATATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGAGG 331
DB 361 ACAAGAAAGGCTTATCTTTACAAACTCGGTGTTATCAATTTGCTGATTGACATGGCAAG 420
QY 332 AATCCGGACCAATCGCTTCGTCGGCGGAGAACTGCTCGGAGCTGGATGGAAC 391
DB 421 AGTTTCAAAGGACCAAGCTTGGTGTCTCAGAACTGCTCTGCCACTTTAAAGGCGACCC 480
QY 392 ACCGGTTTGTGATGGCGTGTCTTCTGTAAGAGGATTTGGAGGAGCAAGGATAGTGA 451
DB 481 ACAAGGTCACAGAACAGCTCTTCTCTGAAACAAAGAGCTGGAGAGAGATGTTATCGTTA 540
QY 452 GCCCTGTAAGAGCAAGAGAGCTGTGGATCTTGTGCTGCACTTTAGTACTACTGAGAC 511
DB 541 GTCCGGTCAAAGATCAGGAGGTTGTGGATCTTGTGGACATTCAGCACTTGGAGCTC 600
QY 512 TAGAGGCTGCATATACACAGCTTAACCTGGAAGAGACACATATTATCTGAAACAGCAACTTG 571
DB 601 TTGAGGAGCTTACCATCAGGATTTGGAAGAGNATATCTCTCTGAGCAACAGCTTG 660
QY 572 TGGACTGTCCCTCAGCATTTCAATTAATTTGGATGCAATGGAGGTTTGTCTTCCCAAGCT 631
DB 661 TGGATTGTCTGGAGCTTTCAATTAATTTGATGTTGCAATGGTGGCTTCTCTTCAAGCT 720
QY 632 TTGAATAGTTTAAGTACATTCGAGGATTCGACAGACAGACAGCTTTATCCATACCTTGTG 691
DB 721 TTGAATACATAAAATCCAAACGGTGGCTCGACACAGAGAAAGCTTTATCTTATACCGGTA 780
QY 692 TCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCAATGATTGATGATA 751
DB 781 AAGATGAAACCTGCAAAATTTTCAGCTGAAAGAGTTGGTGTACAGTCTCTCAACTCAGTCA 840
QY 752 ACATCAACCTGGTGTCTGAGGATGAGTGTGAAGATGAGTGGCTTGGTGGTCCAGTTA 811
DB 841 ACATTAATCTGGGTGCTGAAGATGAACGTAAGATGAGTGGTGGTGGTGGTGGTGGTGGT 900
QY 812 GGGTTGCAATTTGAGGTTGTGAAAGGTTTCAATCTGTACAGAAAGGTTTATATACAGCTG 871
```


Db 901 GCATAGCATTTGAGGTTATACACTGTTCCGGCTTTACAGAGTGGAGTTTACACTGATA 960
Qy 872 ACACCTGTGGAAGAGATCCAATGGATGTGAACCAACGCGAGTTCTTCCCGTCGGTTATGGAG 931
Db 961 GTCACCTGTGGAAGTACTCCAATGGATGTGAACCAACGCGGTTTTCGCCGTTGGTTATGGAG 1020
Qy 932 TCGAGGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGGTGACA 991
Db 1021 TTGAAGACGGGTGTACCATATTGGCTTATTGAAGAACTCATGGGAGCGGATTTGGGGCGACA 1080
Qy 992 ATGGCTACTTTAAGATGGAACCTCGGCAAGAACATGTGTGGTGTGCAACTTGGCGCATCTT 1051
Db 1081 AAGGTTACTTCAAGATGAGATGGGAGAGACATGTGTGGTATTGCTACATGTGCAATCAT 1140
Qy 1052 ATCCCATTTGGCT 1065
Db 1141 ACCCGGTTGGCT 1154

RESULT 7
AAC48945
ID AAC48945 standard; DNA; 1370 BP.
XX AC
AAC48945;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59352.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140895P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

Db 1080 AAGTTACTTCAAGTGGAGATGGGGAACATGTGTGGTATTGCTACATGTGCATCAT 1139
QY 1052 ATCCATTGGGCT 1065
| | | | | | | | | |
Db 1140 ACCCGTTGGGCT 1153
RESULT 8
AAC40628
ID AAC40628 standard; DNA; 1363 BP.
AC AAC40628;
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 28994.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140595P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 02-AUG-1999; 99US-0147038P.
PR 03-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147432P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.

XX 21-NOV-1996 (first entry)
 DT Up-regulated senescence clone, SENU3.
 DE Senescence related gene; expression; inhibition; acceleration; tomato;
 DE lettuce; cabbage; banana, strawberry; wheat; maize; rice; rape; soybean;
 KW delayed senescence; yield; protein content; quality; tolerance;
 KW increased senescence; desiccant; ss.
 XX Lycopersicon esculentum.
 OS
 XX WO9507993-A1.
 PN
 XX 23-MAR-1995.
 XX 13-SEP-1994; 94WO-GB001990.
 XX 13-SEP-1993; 93GB-00018927.
 PR
 XX (ZENE) ZENECA LTD.
 PA
 XX Smart CW, Thomas H, Hosken SE, Schuch WW, Drake CR, Grierson D;
 PI Parrell A, John I;
 PI WPI; 1995-131361/17.
 DR
 XX DNA constructs which modify expression of senescence-related genes -
 PT useful to accelerate or inhibit senescence in crop plants.
 XX
 XX Claim 7; Page 47-48; 78pp; English.
 PS
 XX The sequences given in AAT35095-133 are senescence related genes and
 CC fragments which were used in the construct of the invention which
 CC modifies the expression of at least one senescence related gene. Using
 CC these constructs senescence may be inhibited or accelerated in plants
 CC including tomato, lettuce, cabbage, banana, strawberry, wheat, maize,
 CC rice, rape or soybean. Delayed senescence may indirectly prolong the life
 CC of the plant, increase yield, increase protein content of fruits, improve
 CC quality of leafy vegetables, improve tolerance to disease, drought or
 CC other stress. Increased senescence may more rapidly break down unwanted
 CC plant material and so avoid the use of desiccants on crops. This sequence
 CC represents the full length sequence of the senescence up-regulating
 CC clone, pSENU3, which is also known as 77S3. pSENU3 is a cDNA of 1.1982 kb
 CC encoding a mRNA of approx. 1.4 kb. The mRNA encoded by pSENU3 is
 CC expressed during the onset of senescence in tomato leaves. pSENU3
 CC exhibits 70% homology with oryzain gamma, a cysteine protease expressed
 CC in rice seeds and induced by gibberellin, GA3. pSENU3 is deposited as
 CC NCIMB 40573
 XX
 XX Sequence 1388 BP; 394 A; 268 C; 329 G; 397 T; 0 U; 0 Other;
 SQ
 Query Match 44.4%; Score 475.8; DB 2; Length 1388;
 Best Local Similarity 68.5%; Pred. No. 4.8e-133;
 Matches 657; Conservative 0; Mismatches 302; Indels 0; Gaps 0;
 QY 106 CTTGATCGATGAGCCTGCCATCTCTCGCGCTCTTGGGAGTTGGCCGACGCTTCCAC 165
 DB 125 CCGGATGAGTGGAGAACGGGATCTTCAAGTCGTGCGCCAGACTCGCAGTCTCTCC 184
 QY 166 TTGCGACGGTTGCGCCGAGGTACGGAGAGCTACGGATCGGAGGAGATCAAGAAG 225
 DB 185 TTGCTGCTTTGCTATAGGCAATCGGAAAGGTATGATCTCGTTGAAGAGATCAAGCA 244
 QY 226 AGGTTCGGGATCTTCGTGGAGATCTAGCGTTTATCCGGTCCACTAATCGGAAGATCTG 285
 DB 245 AGGTTTGAGATATTTTGGACAACTGGAAGATGATCCGATCGCATAAACAGAAAGGACTA 304
 QY 286 TCGTATACCTAGGATCAACCAATTCGGCGACCTGACCTGGGAGGAAATTCGGGACCAAT 345
 DB 305 TCATACAACTCGGTATCAATGAGTTTACCGACCTTAACATGGGATGAGTTCGGTAAACAC 364
 QY 346 CGCCTTGGTGGCGGAGAACTCTCGCGGAGTCTCGCGACTGTGCGATGGAACACCGCGTTTGTGAT 405

365 AAGTTGGGGGCATCTCAAAACCTGTTCTGCCACCAAAAGGGAAATCTCAAGCTCACTAAC 424
 QY 406 GCGGTCTCTTCTGTAAACGAGGAGATTGGAGGGAGCAAGGATAGTGAGCCCTGTAAAGGAC 465
 DB 425 GTGTTCTGCCAGAGACGAAAGACTGGAGGAAAGATGGTATTGTTAGCCAGTGAAGGCA 484
 QY 466 CAAGGAAGCTGTGGATCTTCTGCGACTTTCAGTACTACTGAGCACTAGAGGCTGCATAT 525
 DB 485 CAGGGCAAGTGGGGATCTTCTGCTGGACATTCAGCACTACTGGTGCCACTAGAGGACGCTAT 544
 QY 526 ACACAGCTAACTGGAAAGAGCACATCATTTATCTGAACAGCAAACTTTGTGAGACTGTGCTCA 585
 DB 545 GCCCAAGCATTTGGGAAGGGAATCTCTCTGTGAGAGCAGCAGCTTGTGACTGTGCTGA 604
 QY 586 GCATTCAATAACTTTGGATGCAATGAGGTTTGCCTTCCCAAGCCCTTCCAAACGTTAAG 645
 DB 605 GCTTTTAATAACTTTGGTTGCAATGAGGTTTGCCTTCTCAAGCACTTTCAGTACATATAA 664
 QY 646 TACAATGGAGGCATCGACACAGAAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGC 705
 DB 665 TTCAATGGTGGTCTTGATATTGGTGTAGCATATCCATACACCGGCAAGATGGCATATGT 724
 QY 706 AACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCATTTGATTCGATAAACAATCACCCCTGGT 765
 DB 725 AAATTCTCACAGCAAAATATTGGTGTCAAAAGTCATCAGTTCTGTCAATATTACCCCTGGT 784
 QY 766 GCTGAGGATGAGTTGAAGCATGCAGTGGGCTTGGTGGCTCCAGTTAGCGTTGCATTTGAG 825
 DB 785 GCTGAATATGACTGAATTAACGCACTGTCATTTGGTTAGGCTGTTGCTGTTGCTTTGAG 844
 QY 826 GTTGTGAAAGGTTTCAATCTGTACAAAGAAAGGTGTATACAGCAGTGCACCTGTGTGAAGA 895
 DB 845 GTGGTAAAGGGTTCAAAACAGTACAAAGAGCGGAGTTTACGCCAGCACTGAAATGTGGCAG 904
 QY 886 GATCCCATGATGTGAACCAAGCAGTCTTTCGGTCTGCTTATGAGTTCGAGGACGGGATT 945
 DB 905 ACTCCCATGAGCAGTAAACCAATGCTGTCTGTGCGGTTACGGTGTGAAAATGCTACT 964
 QY 946 CCTTATTGGTCTCATCAAGAACTCATGGGGTCAAAATTTGGGGTGCACAAATGGCTACTTTAAG 1005
 DB 965 CCCTACTGGCTCATAAAGAACTCATGGGAGCAGATTGGGGTGGAGATGGATACTTTCAA 1024
 QY 1006 ATGGAATCGGCAAGACATGTGTGTGTGCAACTTGGCATCTTTATCCCATTTGGGC 1064
 DB 1025 ATGGAGATGGGAAAGAAATATGTGTGTGTGCAACTTGTGTGCACTTGTGCACTTACCCATCGTTGC 1083
 RESULT 10
 AAC51435
 ID AAC51435 standard; DNA; 1077 BP.
 XX AC AAC51435;
 XX DT 18-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 68510.
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 XX EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-00301439.
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131444P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138647P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.

PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX Goxlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kriker M, Slater T, Davis KB, Allen K, Hoffman N;
PI Hurban P;
XX WPI; 2002-403163/43.
DR
XX
XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,
PT producing compositions that modulate the expression or function of its
PT encoded protein, and mapping functional regions of protein.
XX
XX Claim 1; SEQ ID NO 20; 49pp + Sequence Listing; English.
XX
XX The invention relates to an Arabidopsis thaliana nucleic acid (I)
CC comprising a sequence capable of hybridising under stringent conditions
CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
CC given in the specification or its fragment. A polypeptide (II) encoded by
CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
CC useful for screening a candidate agent for its biological effect. (I) is
CC useful in identifying homologous or related genes, in producing
CC compositions that modulate the expression or function of its encoded
CC protein, mapping functional regions of the protein and in studying
CC associated physiological pathways. (I) is also useful for the genetic
CC manipulation of cells, particularly plant cells. (I) is also useful in
CC screening assays of various plant strains to determine the strains that
CC are best capable of withstanding a particular disease or environmental
CC stress. (II) and (III) are useful for screening of biologically active
CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
CC pathways. The screened agents are useful in improved methods of treating
CC crops to prevent or treat disease. (II) are also useful in screening
CC programs to identify agents that mimic or enhance the action of tolerance
CC factors. Such agents are useful in improved methods of treating crops to
CC enhance their tolerance to environmental stress. (I) is also useful for
CC enhancing or inhibiting production of a biosynthetic product in a plant.
CC (III) is useful for identifying other mediators that may induce
CC expression of proteins of interest, for establishing the extent to which
CC any specific insect and/or pathogen is responsible for damage to a
CC particular plant, for identifying other mediators that enhance or induce
CC tolerance to environmental stress, for identifying factors involved in
CC biosynthetic pathways of nutritional, commercial, or medicinal value and
CC for identifying productions of nutritional, commercial or medicinal
CC value. (IV) is useful in the study of genetic function and regulation,
CC for alteration of the cellular metabolism and for screening compounds
CC that may affect the biological function of the gene or gene products.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocId=999909770445
XX
SQ Sequence 1282 BP; 367 A; 314 C; 234 G; 367 T; 0 U; 0 Other;

Query Match 43.7%; Score 467.6; DB 6; Length 1282;
Best Local Similarity 66.7%; Pred. No. 1.4e-130;
Matches 668; Conservative 0; Mismatches 334; Indels 0; Gaps 0;

QY 63 CGGTTTCGAAGACCAATCCAAATCCGTCGGTTACACAAAGGCGCTGACTCGAFTGAGCC 122
DB 1214 CGGATTCGATGAATCTAATCAATCAAAATGGTCTCCGACAAATCTCCATGAGCTTGAAGA 1155

QY 123 TGCATCTCCGGGTCCTTGGCAGTTGCGGCACGCGCTTCCACTTCGACGGTTCGCCCG 182
DB 1154 CACCGTCGTCAGATCTCTAGGACAGTCCCGTCATGTTCTATCTCTTTTCGCTTCACTCA 1095

QY 183 CAGGTACCGGAGAGTACAGGATCGGAGGAGATCAAGAAGAGGTTCCGGATCTTCGT 242
DB 1094 CAGGTATCGGAGAGATGATCAGATGTCGAGAGATGAGTAAAGATCTCTGTTTAA 1035

QY 243 CGAGAAATCTAGCGTTTATCCGGTCCACTTAATCGGAAGATCTGCGTATACCCCTAGGAAT 302
DB 1034 GGAGAAATCTTGATTTAAATCAGATCCACTTAATAAGAAAGGCTTATCCCTATAAATCTCTCT 975

QY 303 CAACCAATTCGCGACCTGACCTGGAGGAATTCGGACCAATCCCTTGGTGGCGCGCA 362
DB 974 TAATCAGTTTGTCTGACTTTGACATGGCAAGAGTTTCAAGAGATACAAGCTTTGGAGCTGCTCA 915
QY 363 GAACCTGCTCGGCGACTCGGCATGGAAACACACCGGTTTGTGCGATGGCGTCTTCCCTGTAAAC 422
DB 914 AAATCTGTTCTGCTACTTTAAAGGGTAGCCACGAGATCACTGAAGCTACAGTTCCAGACAC 855
QY 423 GAGGATTCGAGGGAGAGGATAGTAGAGCCTCTTAAGAGCAACGAAGCTGTGGATC 482
DB 854 AAAGATTCGAGAGAGATGGTATTTAGTGGCTCTGTAAGAAACAGGACATTTGGGATC 795
QY 483 TTGCTGGACTTTCACTACTCTGAGGACCTAGAGGCTGCATATACACAGCTAACTTGGAAA 542
DB 794 TTGTTGGACATTTGACACAACTGGAGCTCTTGAACAGCTTACCATCAAGCATTTGGAAA 735
QY 543 GAGCATCATTTATCTGAACAGCAACTTTGTGAGCTGTGCTCAGCATTAATCAATCTTTGG 602
DB 734 AGGAATATCTTTGTCGAGCAACAGCTTTGGATGTTGCTGTACTTTCAATAAATCTTTGG 675
QY 603 ATGCAATGAGGTTTGCCTTCCCAAGCCTTTGAATACGTTAGTACATCAATGGAGGATCGA 662
DB 674 TTGTCATGTTGAGCTTCTCTCAAGCCTTTGAATACATTAATAACAAAGGTTGGGCTCGA 615
QY 663 CACAGAACAGACTTATCCATACCTTTGGTGTCAATGCTATCTGCAACTTCAAGCAGAGAA 722
DB 614 CAGGAGGAGGCTTATCTTACCCGAAAGACGCTGGCTGCAATTTTCAGCGAAAAA 555
QY 723 TGTGTTGTCAGAGTCATTTGATTCGATTAACATCACTACCTCGGTGCTGAGGATGATTGAA 782
DB 554 CATCGGTGTACAGTCCCGTACTCTGTCAACATTAACCTCCGTTGCGAGAGATGAAGTAA 495
QY 783 GCATCAGTGGGCTTGGTGGCTCCAGTTAGCTTGCATTTGAGGTTGTGAAAAGTTTCAA 842
DB 494 GACGCGGTGCGGTTGGTGAAGCCAGTGAAGTGTGCGGTTTGAGGTTGTACATGAATTCAG 435
QY 843 TCTGTACAAAGAGTGTATACAGCAGTGAACCTGTGTGAAGAGATCCAAATGGATGTGAA 902
DB 434 GTTTTATGAAGAGGAGTTTCTAGCAATACATGTGTGTACACTCCCAATGGATGTAA 375
QY 903 CCACGCGTTCTTCGCGTGGTTATGGAGTCGAGGACGCGGATTCTTATTGGCTCATCAA 962
DB 374 CCATGAGGTTTAGCAGTTGGTTATGGAGTTGAACGACGCTCCGCTACTGCTGTATAAA 315
QY 963 GAACCTCATGGGTCACAAATTTGGGGTGACAAATGGCTACTTTAAGATGGAACTCGCAAGAA 1022
DB 314 GAACCTCATGGGAGGTGAATGGGAGACAAATGGCTACTTTCAAGATGGAATGGAAAGAA 255
QY 1023 CATGTGTGTTGTCMAACTTGGCGATCTTATCCCATTTGGC 1064
DB 254 CATGTGTGTTGTCMAACATGTTTCATCGTATCCCGTTGTAGC 213

RESULT 12
AAC44535
ID AAC44535 standard; DNA; 1072 BP.
XX
XX AAC44535;
AC AC
XX XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 43187.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic; pathway;
KW promoter; termination sequence; corn; ss.
XX
XX Zea mays subsp. mays.
XX
XX EP1033405-A2.
FN
PD 06-SEP-2000.

XX	25-FEB-2000; 2000EP-00301439.	PR	08-JUL-1999;	99US-0142803P.
PF		PR	09-JUL-1999;	99US-0142920P.
XX		PR	12-JUL-1999;	99US-0142977P.
PR	25-FEB-1999;	PR	13-JUL-1999;	99US-0143542P.
PR	05-MAR-1999;	PR	14-JUL-1999;	99US-0143624P.
PR	09-MAR-1999;	PR	15-JUL-1999;	99US-0144005P.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144085P.
PR	25-MAR-1999;	PR	16-JUL-1999;	99US-0144086P.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144325P.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144331P.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144332P.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144333P.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144334P.
PR	19-APR-1999;	PR	19-JUL-1999;	99US-0144335P.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144352P.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144632P.
PR	28-APR-1999;	PR	20-JUL-1999;	99US-0144884P.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0144814P.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086P.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087P.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145089P.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192P.
PR	06-MAY-1999;	PR	23-JUL-1999;	99US-0145145P.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145218P.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919P.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386P.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388P.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389P.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038P.
PR	27-MAY-1999;	PR	03-AUG-1999;	99US-0147204P.
PR	01-JUN-1999;	PR	04-AUG-1999;	99US-0147302P.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147192P.
PR	04-JUN-1999;	PR	05-AUG-1999;	99US-0147260P.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147303P.
PR	08-JUN-1999;	PR	06-AUG-1999;	99US-0147416P.
PR	10-JUN-1999;	PR	08-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935P.
PR	14-JUN-1999;	PR	10-AUG-1999;	99US-0148171P.
PR	16-JUN-1999;	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149302P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149302P.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151080P.
PR	18-JUN-1999;	PR	30-AUG-1999;	99US-0151303P.
PR	18-JUN-1999;	PR	31-AUG-1999;	99US-0151438P.
PR	21-JUN-1999;	PR	01-SEP-1999;	99US-0151330P.
PR	22-JUN-1999;	PR	07-SEP-1999;	99US-0152363P.
PR	23-JUN-1999;	PR	10-SEP-1999;	99US-0153070P.
PR	24-JUN-1999;	PR	13-SEP-1999;	99US-0153758P.
PR	28-JUN-1999;	PR	15-SEP-1999;	99US-0154018P.
PR	29-JUN-1999;	PR	16-SEP-1999;	99US-0154039P.
PR	30-JUN-1999;	PR	20-SEP-1999;	99US-0154779P.
PR	01-JUL-1999;	PR	22-SEP-1999;	99US-0155139P.
PR	01-JUL-1999;	PR	23-SEP-1999;	99US-0155486P.
PR	02-JUL-1999;			
PR	06-JUL-1999;			

PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135253P.
PR 24-MAY-1999; 99US-0135623P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137503P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139113P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 08-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151130P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Query Match 40.1%; Score 429.4; DB 3; Length 2149;
Best Local Similarity 71.1%; Pred. No. 6.7e-119;
Matches 582; Conservative 0; Mismatches 236; Indels 1; Gaps 1;
QY 247 AATCTAGCGTTTATCCGCTCCACTAATCGGAAGATCTGTGTATACCTAGGAAATCAAC 306
DB 1113 AATCTTGAATTTGATCAGATCCACCAAGAAAGCTTATCTTACAAACTCGGTGTTAAT 1172
QY 307 CAATTCGCGGACCTGACCTGGGAGGAATTCGGACCAATCCCTTGTGGCGGCAGAAC 365
DB 1173 CAATTTGCTGATTGACATGCAAGAGTTTCAAGGACCAAGCTTGTGCTGTCTCAGAAC 1232
QY 367 TGCTCGGCGACTGCGCATGGAACACACCGGTTTGTGATGCGGTCTCTCTGTAACGAGG 426
DB 1233 TGCTCTGCCACTTTAAAGGGCAGCCACCAAGGTCACAGAAGCAGCTCTTCTGAAACAAA 1292
QY 427 GATTGGGGGAGCAAGGATAGTGGCCCTGTAAGGACCAAGGAGCTGTGGATCTTGC 486
DB 1293 GACTGGAGAGATTTGATCGTTAGTCCGTCGAAGATCAGGAGGTGTGGATCTTGC 1352
QY 487 TGGACTTTCACTACTTGGAGCACTAGAGCTGCATATACACAGCTAACTGGAAGAGC 546
DB 1353 TGGACATTCAGCAGCTTGGAGCTTGGAGCAGCTTACCATCAGGCAATTTGGAAAGGA 1412
QY 547 ACATCATTTCTGACACCACTTGGGACTGTGCTCAGCATTAATTAATCTTTGGATGC 606
DB 1413 ATAATCTCTCTGAGCAACAGCTTGTGATTTGTGAGCTTCAATTAATTAATTTGGTGC 1472
QY 607 AATGAGGTTTGCCTTCCCAAGCCCTTTGAATACGTTAAGTAACTGAGGAGCATCGACAC 666
DB 1473 AATGTTGCTTCTTCTTCAAGCCCTTTGAATACATAAATCAACGGTGGCTCGACACA 1532
QY 667 GAACAGACTTATCCATCTTGGTGTCAATGGTATCTGCACTTCAAGCAGGAGAAATGT 726
DB 1533 GAGAAAGCTTATCTTATACCGGTAAGATGAAACCTGCAAAATTTTCAGCTGAAAGCGT 1592
QY 727 GGTGTCAAGGTCATTGATTCGATAAATCAATCACTTGGTGTGAGGATGATTTCAAGCAT 786
DB 1593 GGTGTCAAGTCTTCACTCACTCACTTACTCTGGTGTGAGGATGATTTCAAGCAT 1652
QY 787 GCAGTGGCTTGGTGGCTTCCAGTTAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 846
DB 1653 GCGTTTGGATTGGTACGGCCAGTAGCATAGCATTTGAGGTTATACACTGTTCCGCTT 1712
QY 847 TACAAGAAAGTGTATACAGCAGTGCACACCTGTGGAAGAGATCCAATGGATGTCAACAC 906
DB 1713 TACAAGAGTGGATTTACATGATAGTCACTGTGGAAGTATCCANTGGATGTGAACAC 1772
QY 907 GCAGTTCTTGGCTGGCTTATGAGTGCAGGACCGGATTCCTTATTTGGCTCATCAAGAAC 966
DB 1773 GCGTTTGGCTGGCTTATGAGTGTCAAGCAGCTGTACCATATTTGGCTTATTAAGAAC 1832
QY 967 TCATCGGCTCAAAATTTGGGTTGCAAACTGCTTAAATGGAAGTCTGCGCAAGCATG 1026
DB 1833 TCATCGGAGCGGATTTGGGGGACAAAGGTTTACTTCAAGATGGAGATGGAGATGGAGAT 1891
QY 1027 TGTGTGTTGCAACTTTGGGCTATCTTATCCCATTTGGCT 1065
DB 1892 TGTGTGTTGCTTACATGTGATCATCATACCCCGTTGGCT 1930

RESULT 14

ADB94877

ID ADB94877 standard; cdna; 1652 BP.

XX

AC ADB94877;

XX

DT 04-DEC-2003 (first entry)
XX Programmed cell death pathway cysteine protease cdna #42.
DE
XX
XX programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
KW Dnase; lls; lsd1; nucellin-like aspartic protease; annexin; prohibitin;
KW fen-like protein; rac2; retinoblastoma-related protein; SINA; TFIID;
KW TEGT; xylogenic RNase; pur-alpha; cytochrome protease; RPP5-like protein;
KW GP 91 NADPH oxidase subunit; NPR-like protein; BAG-1;
KW defender against cell death; lethal leaf spot; lesion stimulating death;
KW seven in absentia; transcription initiation factor;
KW testis enhanced gene transcript; gene; ss.
OS Pinus radiata.
XX
XX US2003082724-A1.
PN
XX
XX 01-MAY-2003.
PD
XX
XX 14-AUG-2002; 2002US-00219220.
PF
XX
XX 04-JUN-1999; 99US-00325932.
PR (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Flinn B, Lasham A;
PI
XX
XX WPI; 2003-786916/74.
DR P-PSDB; ADB94919.
DR
XX
XX New isolated polynucleotide useful for modulating programmed cell death,
PT altering the development cycle of plant cells, and subsequently modifying
PT plant development.
XX
PS Claim 1; Page 154; 214pp; English.
XX
XX The invention describes an isolated polynucleotide (I) comprising a
CC nucleotide sequence that is one of 145 fully defined sequences of 221-
CC 3415 base pairs (bp), given in the specification, and/or its complements,
CC reverse complements, reverse sequences, or sequences having 75, 90 or 95
CC % sequence identity to or that hybridise under stringent hybridisation
CC conditions to one of the 145 sequences. The methods and compositions of
CC the present invention to do with (I) are useful for modulating programmed
CC cell death and thereby altering the development cycle of plant cells, and
CC altering plant development. This sequence encodes a protein associated
CC with the programmed cell death pathway.

SQ Sequence 1652 BP; 476 A; 318 C; 393 G; 465 T; 0 U; 0 Other;

Query Match 38.0%; Score 407.4; DB 9; Length 1652;
Best Local Similarity 65.7%; Pred. No. 2.7e-112;
Matches 626; Conservative 0; Mismatches 321; Indels 6; Gaps 2;

QY 116 TTGAGCTGCCATCTCTGGGCTCTTGGCAGTTGGCGCCAGCTTCCACTTCGACCGT 175
DB 385 TGGAGTCCAAATTCGTTAAATCTTGGGAACCAATCCCAAGTCTTACAAATTCGGAGT 444
QY 176 TCGCCCGCAGGTACCGGAAAGAGCTACGATCGGAGGAGGAGATCAAGAAGAGTTCCGGA 235
DB 445 TCCTGTGAGATATGCAAGAGTACGATTCGTCCATCAGCTTGTGCATAGATTCAATG 504
QY 236 TCTTGTGAGAAATCTAGGTTTATCCGTTCACTAATCGGAAGGATCTGTGATATACC 295
DB 505 TCTTTGTGAAGAACGTTGGAGCTGATCGAGTCAAGAAACAGAATGAAGTTCCTTACTT 564
QY 296 TAGGAATCAACCAATTCGCGGACCTGACCTGGGAGGAATTCGCGACCAATTCGCTTGGTG 355
DB 565 TGCAATTAATGATTTGCTGACATAACATGGGAGGAATTCATGACAAATATTCGGTG 624
QY 356 CGCGCAGAACTGCTCGGAGCATGCGCATGGAAACCAACCGGTTTGTGATGGGTCTTC 415
DB 625 CTTTACAGAACTGTTCCG---CTACCCACAGTAACCAATAAGTTGACGTATGCCAGCTTC 681

QY 416 CTGTAAACGAGGATTTGGAGGGAGCAAGGATAGTGTAGCCCTGTAAAGACCAAGGAGCT 475
Db 682 CTCGAAAGAAAGACTGGAGCAAGGACCTAGTGTAGTCTGTAAAGACCAAGGAGCT 741
QY 476 GTGGATCTTGTGGACTTTTCACTACTACTGAGGAGCAAGGAGCTGTATACACAGCTAA 535
Db 742 GTGGATCTTGTGGACTTTTCACTACTACTGAGGAGCAAGGAGCTGTATACACAGCTAA 801
QY 536 GTGGAAAGAGCAATCATATCTGAACAGCACTTGTGTGAGCTGTATACACAGCTAA 595
Db 802 CAGGAAAGAGCACTTATCTGTCTGAACAGCACTTGTGTGAGCTGTATACACAGCTAA 861
QY 596 ACTTTGGATGCAATGAGGAGTTTCCCTTCCCAAGCCTTTGAATACGTTAAAGTACATGAG 655
Db 862 ACTTTGGATGCAATGAGGAGTTTCCCTTCCCAAGCCTTTGAATACGTTAAAGTACATGAG 921
QY 656 GCATGACACAGACAGACTTATCCATACCTTTGGTGTCAATGGTATCTGCAACTTCAAGC 715
Db 922 GCGTTGACACTGAGGAGGCTTATCTTACACGCTTAAAGATGCTTGTGTTAAATATGATG 981
QY 716 AGGAGATCTTGTGTCAAGGCTTATGATTCGATTAACATACCTGCTGAGGATG 775
Db 982 TAAACAATCTCGGTGTAAAGGTTGCTGATGTCATCAACATCAGTTTGGGGCAGAGATG 1041
QY 776 AGTTGAAGCATGAGTGGGCTTGTGCTGCTCCAGTTAGCTTCCATTTGAGGTTGTGAAG 835
Db 1042 ACCTTCAAGCTGCTGTGGGTAGTTTGGCTGTGCTAGTGTGCAATTCACAGTTAATTCGG 1101
QY 836 GTTTCATCTGTACAAAGAGGTGTATACAGCAGTGTGACACTGTGGAAGAGATCCAAATGG 895
Db 1102 AATTTCCGATTTTCAAGGAGGAGCTTTTCAAGCAAGATGTTGTGTCAAGGCCCAATGG 1161
QY 896 ATGTGAACACAGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 952
Db 1162 ATGTCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1221
QY 953 GCGTCAATCAAGAACTCATGCGGTGTCAAAATGGGTGACAAATGGGTGACATTTTATGATGGA 1012
Db 1222 GCGTCAATCAAGAACTCATGCGGTGTCAAAATGGGTGACAAATGGGTGACATTTTATGATGGA 1281
QY 1013 TCGGCAAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1065
Db 1282 TAGGGAAGATATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1334

RESULT 15
ADB94874
ID ADB94874 standard; cDNA; 888 BP.
XX AC ADB94874;
XX DT 04-DEC-2003 (first entry)
XX DE Programmed cell death pathway cysteine protease cDNA #39.
XX KW Programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
KW Dnae; lls; lsd; nucellin-like aspartic protease; annexin; prohibitin;
KW fen-like protein; rac2; retinoblastoma-related protein; SINA; RFIID;
KW TEGT; xylogenic kinase; pur-alpha; cytochrome protease; RPP5-like protein;
KW sp 91 NADPH oxidase subunit; NPR-like protein; BAG-1;
KW defender against cell death; lethal leaf spot; lesion stimulating death;
KW seven in absentia; transcription initiation factor;
KW testis enhanced gene transcript; gene; ss.
XX OS Eucalyptus grandis.
XX FN US2003082724-A1.
XX PD 01-MAY-2003.
XX PF 14-AUG-2002; 2002US-00219220.
XX PR 04-JUN-1999; 99US-00325932.

XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Flinn B, Lasham A;
XX WPI; 2003-786916/74.
XX P-PSDB; ADB94916.
XX New isolated polynucleotide useful for modulating programmed cell death,
XX altering the development cycle of plant cells, and subsequently modifying
XX plant development.
XX Claim 1; Page 152; 214pp; English.
XX The invention describes an isolated polynucleotide (I) comprising a
XX nucleotide sequence that is one of 145 fully defined sequences of 221-
XX 3415 base pairs (bp), given in the specification, and/or its complements,
XX reverse complements, reverse sequences, or sequences having 75, 90 or 95
XX % sequence identity to or that hybridise under stringent hybridisations
XX conditions to one of the 145 sequences. The methods and compositions of
XX the present invention to do with (I) are useful for modulating programmed
XX cell death and thereby altering the development cycle of plant cells, and
XX altering plant development. This sequence encodes a protein associated
XX with the programmed cell death pathway.
XX Sequence 888 BP; 205 A; 240 C; 249 G; 194 T; 0 U; 0 Other;
XX
XX Query Match 32.8%; Score 351.8; DB.9; Length 888;
XX Best Local Similarity 66.9%; Pred. No. 1.3e-95;
XX Matches 500; Conservative 0; Mismatches 247; Indels 0; Gaps 0;
QY 116 TTGAGCTGTCATCTCGGCTCTTGGAGTTGCGGCACTTCCGCTTCCACTTCGACGGT 175
Db 142 TTGAGTCTCTCATCTCTCAACCGTCTGCGCCACACCGCTCCGCCCTCTCTTCTGACT 201
QY 176 TCGCCCGCAGTACCGGAGAGGATACGATCGGAGGAGGAGATCAAGAGAGGTTCCGGA 235
Db 202 TCGCCCGGCGCAGGAGAGGATTAAGAGAGCGGAGGAGATCAAGTTGAGTTGATA 261
QY 236 TCTTGTGGAGATCTAGCGTTTATCCGCTCCACTATCGAGGAGATCTGCTGTATACC 295
Db 262 ATTACAGGAGATCTCAAGCTCATTCGATCCCAACCAAGAGAGGCTTGCCTTACTC 321
QY 296 TAGGAATCAACCAATTCGCGACCTGACCTGGGAGGAGATTCGGACCAATCGCTTGGT 355
Db 322 TCGTGTATCATGATGCTGACTGGAGCTGGGAGGAGTTCAAGCCACAGACTGGAG 381
QY 356 CGGCGCAGAACTGCTCGGCGACTCGGCATGGAACACCGGTTTGTGATGGCGTGTTC 415
Db 382 CTTCCTAAGACTGCTCTGCCACCAAGGAGGAGGAGCTCAAGCTCAGGAGATTTCTTC 441
QY 416 CTGTAAACGAGGATTTGGAGGGAGCAAGGATAGTGTAGCCCTGTAAAGACCAAGGAGCT 475
Db 442 CCGAAACGAAAGACTGGAGAGAAAGGCAATTTAAGCCCAAGTTAAGATCAAGCGGCT 501
QY 476 GTGGATCTTGTGGACTTTTCACTACTACTGAGGAGCTGTATACACAGCTAA 535
Db 502 GTGGATCTTGTGGACTTTTCACTACTACTGAGGAGCTGTATACACAGCTAA 561
QY 536 CTGGAAGAGCAGATCATATCTGAACAGCACTTGTGAGCTGTGCTTCAATA 595
Db 562 ACAGGAAAGGAAATCTCTGTCTGAGCAGCAGCTGCTGAGCTGCTTCAACA 621
QY 596 ACTTTGGATGCAATGAGGAGTTTGGCTTCCCAAGCCTTTGAATACGTTAAAGTACATGAG 655
Db 622 ACTTTGGATGCGGTGGGCTTGGCTTGGAGGAGGAGTTCGAGTACATCAAGTACACGCTG 681
QY 656 GCATGACACAGACAGACTTATCCATACCTTGTGTCATGCTGCAATGGTATCTGCAACTTCAAGC 715
Db 682 GCGTTGAGACCGGAGGAGCTTATCTTACATGTCAGCAAGAAATGGTACCTGCAATTTCTCG 741
QY 716 AGGAGATGTTGGTGTCAAGGCTTATGATTCGATTAACATCAACCTGCTGAGGATG 775

Db 742 CTGGCAAGGTCGCTGTCAAGTTGTGCACTCTGTCAACATCTCTATGGGTGCTGAGGATG 801
Qy 776 AGTTGAAGCATGCAGTGGGCTTGGTCCGTCACGTTAGCGTTGCATTGAGGTTGTGAAG 835
Db 802 AACTTAAAGCATGCAGTTGGCTGGTCCGGCCAGTCAGTGTGGCATTCCAGGTCACGGATG 861
Qy 836 GTTTCATCTGTACAGAAAGGTGTAT 862
Db 862 GCTTCAGCTCTACGAGTCGGGTGTGT 888

Search completed: March 29, 2004, 19:38:48
Job time : 378 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 19:19:12 ; Search time 78 Seconds
(without alignments)
7619.903 Million cell updates/sec

Title: US-10-087-714-1

Perfect score: 1071

Sequence: 1 atggcagctaaagctctctt.....atccattgtgctgtgtag 1071

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfilest.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	541.4	50.6	1407	4	US-09-325-932A-107
2	280.2	26.2	677	4	US-09-325-932A-111
3	205.2	19.2	475	4	US-09-325-932A-120
4	161.4	15.1	1600	4	US-09-325-932A-98
5	160.2	15.0	1474	3	US-08-821-994-64
6	159	14.8	1833	3	US-09-120-365-6
7	159	14.8	1833	3	US-09-515-039-6
8	157.6	14.7	1056	1	US-09-500-651-1
9	157.6	14.7	1056	1	US-08-813-591-1
10	155	14.5	522	4	US-09-325-932A-112
11	150.6	14.1	1366	3	US-08-883-526-2
12	150.6	14.1	1366	4	US-09-519-283-2
13	147.4	13.8	1390	3	US-08-821-994-61
14	147.4	13.8	1441	3	US-08-821-994-63
15	145.8	13.6	1102	3	US-08-821-994-86
16	144.2	13.5	1434	3	US-08-821-994-62
17	139.4	13.0	990	2	US-08-806-959-1
18	137.2	12.8	1661	3	US-08-821-994-82
19	134	12.5	1575	4	US-09-784-642-1
20	131.6	12.3	1553	3	US-08-821-994-60
21	127.6	11.9	1463	4	US-09-325-932A-92
22	125	11.7	1482	1	US-08-330-121B-1
23	125	11.7	1482	1	US-08-536-861-1
24	125	11.7	1482	5	PTC-US95-13820-1
25	125	11.7	1614	4	US-08-684-932A-35
26	125	11.7	1619	1	US-08-208-007A-1
27	125	11.7	1619	4	US-08-915-095A-1
28	125	11.7	1619	4	US-08-798-096-1
29	125	11.7	1619	4	US-08-798-096A-1
30	125	11.7	1619	4	US-09-553-956-1
31	125	11.7	1619	4	US-08-553-125A-1
32	125	11.7	1619	4	US-10-114-464-1
33	125	11.7	1619	5	PTC-US94-04781-1
34	125	11.7	1669	3	US-08-964-308-3
35	125	11.7	1669	3	US-08-964-313-3
36	125	11.7	1669	4	US-09-069-138-3
37	123.4	11.5	990	3	US-08-964-308-12
38	123.4	11.5	990	3	US-08-964-313-12
39	123.4	11.5	990	4	US-09-069-138-12
40	121.8	11.4	990	3	US-08-964-308-13
41	121.8	11.4	990	3	US-08-964-313-13
42	121.8	11.4	990	4	US-09-069-138-13
43	121.8	11.4	1577	3	US-08-821-994-59
44	118.6	11.1	660	3	US-09-004-731-94
45	118.6	11.1	660	3	US-08-749-699-94

RESULT 1

US-09-325-932A-107

; Sequence 107, Application US/09325932A

; Patent No. 6451604

; GENERAL INFORMATION:

; APPLICANT: Flinn, Barry

; APPLICANT: Lasham, Annette

; TITLE OF INVENTION: Compositions affecting programmed cell

; TITLE OF INVENTION: death and their use in the modification of forestry plant devel

; FILE REFERENCE: 1022

; CURRENT APPLICATION NUMBER: US/09/325,932A

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 206

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 107

; LENGTH: 1407

; TYPE: DNA

; ORGANISM: Eucalyptus grandis

US-09-325-932A-107

Query Match 50.6%; Score 541.4; DB 4; Length 1407;

Best Local Similarity 70.5%; Pred. No. 1.8e-157; Indels 3; Gaps 1;

Matches 738; Conservative 0; Mismatches 306;

Qy	21	CTTCCTACTCTTCTCGTCTCGCCCTCTCCGTTCGGCTCGCGGTTTCGAGAACAA	80
Db	87	CGTCCTCTCTCTCGTCTCGCTCTCGCGGCTCGAGCTTCGAGAGTCCAA	146
Qy	81	TCCAAATCCGGTCGGTACACA---AAGGCTGACTGATGAGCTGCCATCTCGGCT	137
Db	147	CCCATTCGGCTCTTCCCGAGCGGCTTCGGAGCTTCATGTCAGAT	206
Qy	138	CTTGGCAGTTGCGGCACGCTTCACCTTCGACCGGTTCCCGCAGTACGGGAAGAG	197
Db	207	CGTCGCGCACCACCGCCCTTCTCTTTCGCGCTTCGCAACAGGTATGGGAAGAG	266
Qy	198	CTACGAGTTCGAGAGGAGATCAAGAGGTTTCGGATTCGAGAACTACGTT	257
Db	267	GTACGAGACCGCGAGGAGATCAAGTCTCGGTTTCGAGATCTTCAGGAGATCTCAAGT	326
Qy	258	TATCCGCTCCACTAATCGGAAGGATCTGCTATACCTAGGAATCAACCAATTCGCGA	317
Db	327	GATCCGATCCACCAACAAGAGGGCTTCCTACACCTCGGTGTCAATAAGTTCTGA	386
Qy	318	CTGAGCTGGAGGAATTCGGAACCAATCGCTTCGTCGGCGGACAACTGCTCGGAC	377
Db	387	TTGAGCTGGAGGAGTTTACGAGGACAGACTGGAGTGTCTCAAACTGCTCTCCAC	446
Qy	378	TGCGCATGGAACACCGGTTTCTGATGCGCTTCTCTGTAACGAGGATTCGAGGA	437

ALIGNMENTS

Db 447 CACCAAGGGCAACCAAGCTCACCAGCAAGCTCTTCCCGAGATGAAAGAGCTGGAGAGA 506
Qy 438 GCAAGGGATAGTAGAGCCCTGTAAAGCAACCAAGAAAGCTGTGATCTTGTGACTTTTCAG 497
Db 507 AAAGGGCATGTAGCCCAATTAAGATCAGGGGCACTGTGATCTTGTGACTTTTCAG 566
Qy 498 TACTACTGAGCACTAGAGGCTGCGATATACAGAGCTAATCTGAAAGAGCAACATCATTC 557
Db 567 TACCACTGGAGCTCTTGGAGCTGCTTATCACAAGCATTCGGGGAACAAATCTCTCTGTC 626
Qy 558 TGAACAGCAACTTGTGAGCTGTGCTCAGCATTTCAATTAATTTGATGCAATGGAGGTTT 617
Db 627 TGAGCAGAGCTCTGAGATCTGTGCTGGGCTTTCAACAACTTGTGATGTAGTGGACT 686
Qy 618 GCCTTCCCAAGCTTTGAATACGTTAAGTAACTAAGTAACTGAGGATCGACACAGAACTTA 677
Db 687 GCCATCCCAAGCTTTGAGTACGTCAAGTACAAAGCTGGCTTGTATCCGAGGAAAGCATA 746
Qy 678 TCCATACCTTGGTCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGT 737
Db 747 TCCTTATACCGCAGTGGATGTAGCTGCAATTTCTCGGCTGATATGTTGGTGTCAAGT 806
Qy 738 CATTTGATTCGATAAATCACTCACCCTGGGTGCTGAGGATGAGTTGAAGCATGCAAGTGGCTT 797
Db 807 GCTCGACTCTGTTAAATCACTCACCCTGGGTGCTGAGGATGAATTAAGCATGCAAGTGGCTT 866
Qy 798 GGTGCTGCTAGCTAGGCTTGCATTTGAGTTGAGTTGAAAGGTTTCAATCTGTACAGAAAG 857
Db 867 CGTCCGGCCAGTGAAGTGTGCGCATTCAGGTCGTGAAAGACTTCAAGATTGTACAGTCGG 926
Qy 858 TGTATACAGCAGTGACACCTGTGGAAGAGATCAATGAGTGAACCAAGCAGCTTCTTGC 917
Db 927 TGCTACACAGCGATACATCGGTAGCATTCCATGATGGAACCATGCTTCTGCTGC 986
Qy 918 CGTGGTATAGAGTCGAGACCGGATTCCTTATTTGGCTCATCAAGAACTCATGGGTAC 977
Db 987 TGTGTTATAGAGTTGAAGATGGTGTCTCGTCTCATCAAGAAATTCCTGGGGAGC 1046
Qy 978 AAATGGGGTGACAATGGCTACTTTAAGATGGAACCTCGGCAAGAAATGTTGGTGTGC 1037
Db 1047 AGACTGGGTGACCAAGGATACCTCAAGATGGAGATGGGAAAGAAATGTTGGAGTCCG 1106
Qy 1038 AACTTGGCATCTTATCCCAATTTGGC 1064
Db 1107 TACTTGGCATCATACCTGTTGGC 1133

RESULT 2
US-09-325-932A-111
; Sequence 111, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 111
; LENGTH: 677
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-111

Query Match 26.2%; Score 280.2; DB 4; Length 677;
Best Local Similarity 68.1%; Pred. No. 9.6e-77;
Matches 390; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 116 TTGAGCTGCGCATCCTCGGCTCTTGGCAGTTTCCGCCACGCTTCCATTCGACGGT 175

Db 102 TCAGTCTCTCCATCTCCAAACCGTTGGCCACGCGCTCCCGCCCTCTCTCTTCTAGACT 161
Qy 176 TCGCCCGCAGGTACGGGAAGAGCTACGGATCGAGAGGAGATCAAGAGAGGTTTCGGA 235
Db 162 TTGCCAGAGCTACGAGAGAGGTACGAGACAGCGGAGGAGATCAAGTTGAGGTTTCGATA 221
Qy 236 TCTTCGTGGAGAACTTAGCGTTTATCCGTTCCACTAATCGGAAGGATCTGTCTGTATACCC 295
Db 222 ATTACAGGAGAACTCTCAAGCTCATTCGATCCACCAACAGAGGCTTGCCTTACACTC 281
Qy 296 TAGGAATCAACCAATTCGCCACTGACCTGGGAGAAATTCGGACCAATCGCTTGGTG 355
Db 282 TCGCTGTTAATCAGTATGCTGACTGGAGCTGGGAGGAGTTCAAGACGCAAGACTGGGAG 341
Qy 356 CGGCGCAGAACTGCTCGGCACTGCGCATGGAACCAACCGGTTTGTGATGGGCTGCTTC 415
Db 342 CTTCTCAAGACTGCTCTGCCACCAACAGGCGAGCCCAAGCTCACGAGCTGTTCTTC 401
Qy 416 CTGTAACGAGGAGTTGAGGAGCAAGGATAGTGAGCCCTGTAAAGGACCAAGAAAGCT 475
Db 402 CCAAAACGAAAGACTGGAGAAAGAGGGCATTTGAGCCCAAGTTAAATAATCAAGGGCGCT 461
Qy 476 GTGATCTTGTGACTTTTCACTACTGAGCACTAGAGGCTGACATATACACAGCTAA 535
Db 462 GTGATCTTGTGAGTTTCAAGCCACTGAGCTCTCGAGGCTGCTTATCACCAAGCAC 521
Qy 536 CTGAAAGAGACATCATTTATCTGAACAGCAACTTGTGACTGTGCTCAGCAATCAATA 595
Db 522 ACGGAAAGGAATCTCTCTGTCGAGCAGCAGCTCGTGGACTGCGCTTTCACCA 581
Qy 596 ACTTTGATGCAATGGAGGTTTGGCTTCCAGAGCTTTGAATAGCTTAACTAGCAATGGAG 655
Db 582 ACTTTGATGCGATGGGGGTTGCCGTCAAGCCTTCGAGTACATCAAGTACACCGTG 641
Qy 656 GCATCGACACAGAACAGACTTATCCATACCTTG 688
Db 642 GCCTTGACCGAGAAAGCTTATCCTTATCTG 674

RESULT 3
US-09-325-932A-120
; Sequence 120, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-325-932A-120

Query Match 19.2%; Score 205.2; DB 4; Length 475;
Best Local Similarity 66.0%; Pred. No. 1.3e-53;
Matches 313; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

Qy 151 CGCCAGCCTTCCACTTCGCACGGTTGCGCGCAGGTACGGGAAGAGCTACGATCGGAG 210
Db 5 CCCAAGTCTCTAGAGTTTGGAGTTCCGTCTCAGATATGCAAGAGAGTACGATCTGTC 64
Qy 211 CAGGAGATCAAGAAAGAGTTTCGGATCTTCTGTGAGAAATCTAGCGTTTATCCGCTCACT 270
Db 65 CATCAGCTTGTGCATAGATTCAATGCTCTTGTGAAGAACGTCGAGCTGATCGAGTCAAGA 124
Qy 271 RATCGAAGGATCTGCTGATACCTAGGAATCAACCAATTCGCGCACTGACCTGGGAG 330

Db 125 AACAGATGAAGCTTCTTATATCTTGGCAATAAATGATTTGCTGACATAAATCATGGAG 184
Qy 331 GAATTCGGACCAATCGCTTGGTGGCGGAGAACTGCTCGGAGCTGCGCATGGAAC 390
Db 185 GAATTCATGACAAATATTTGGTGGCTTCAAGAACTGTTTGG---CTACCCACAGTAAC 241
Qy 391 CACCGGTTTTCGATGGGCTCTTCTGTAACGAGGATTTGGAGGGAGCAAGGATAGTG 450
Db 242 CATAGTTGAGTATGCCAGCTTCTCGNAGAAAGACTGGAGACAAAGAGGCATAGTG 301
Qy 451 AGCCTGTAAAGGACCAAGGAGCTGTGGATCTTGGACTTTTCAGTACTACTTGGAGCA 510
Db 302 AGTCTGTAAAGGACCAAGGAGCTGTGGATCTTGGACTTTCAGCACTTTCAGCACTTCTGGAGCA 361
Qy 511 CTAGAGCTGCATATACAGCACTAATCTGGAAGGACACATCATTTATCTGACAGCACTT 570
Db 362 CTAGAGCTGCATATACAGCACTAATCTGGAAGGAGCTGTTATCTGCTGACAGCACTG 421
Qy 571 GTGAGCTGTGCTCAGCACTTCAATTAATTTGGATGCAATGGAGGTTTGGCTTCC 624
Db 422 GTTGAATGTGCTGGAGCAATTTAACAATTTGGTTGCAATTTGGTGGAGTGGCCATCC 475

RESULT 4
US-09-325-932A-98
; Sequence 98, Application US/09325932A
; Patent No. 6451804
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant develop
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 1600
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-98

Query Match 15.1%; Score 161.4; DB 4; Length 1600;
Best Local Similarity 52.9%; Pred. No. 1e-39;
Matches 460; Conservative 0; Mismatches 376; Indels 33; Gaps 4;
Qy 165 CTTCGACGGTTCCGCCAGGTACGGGAGAGCTAGCGATCGGAGGAGAGATCAAGAA 224
Db 244 CTACAGAGCTGGCTCGCAAGCAGCGCAAGGCTTACACGCCCTTGGCGAGAGAGAA 303
Qy 225 GAGGTTGGGATCTTCGTTGAGAAATCTAGCGTTTATCGGTTCCACTAA---TCGGAAGGA 281
Db 304 GCGCTTCAGGTCTTCAAGGACAACTCCGCTTCATCGACGACCAACGCCGCGGGGA 363
Qy 282 TCTGTCGATACCTTAGGAATCAACCAATTCGCGAGCTGACCTGGGAGGAATTCGGAC 341
Db 364 CGGACCTTACGGTTCGCTTCAACGATTCGCGAGCTTCCGAGCTTCTAACCGAGATACCGGTC 423
Qy 342 CAATGCTCTTGGTGGCGGCGAGAACTGCTCGGCACTGCGCATGGAACCAACCG----- 395
Db 424 CATGTACCTGGCGGCGAGGATGATCGGTTCGGGCGCGGCTCGGAGGGGCGCGCGCA 483
Qy 396 -----GTTTGTGATGGCTGCTTCTGTAACGAGGATTTGGAGGGAGCAAGG 443
Db 484 TCGGTACCGCTGGCGGCGGAGAGCTGCCGCGCTTCGATTTGGAGAGAGGAGG 543
Qy 444 GATAGTGAAGCTGTAAAGGACCAAGGAGCTGTGGATCTTGTGAGCTTTTCACTACTAC 503
Db 544 GCGCGTGTGACGCTCAAGGACCAAGGAGCTGCGGAGTTGCTGGGCGTTTCTCTACAAT 603
Qy 504 TGGAGCACTAGAGGCTGCATATACAGCTAATCTGGAAGAGCAATCATTTATCTGACA 563

Db 604 TGCTGCTGGAGGGGATAAACAAGCTTGTGACTGGTGGTGGTGGTGGTGGTGGTGGTGG 663
Qy 564 GCAACTGTGGAGCTGTGCTCGCTCAGCAATCAATAAATTTGGATGCAATGAGGTTTGCCTTC 623
Db 664 GGAACCTGTGGAGCTGCGATACATCTCTACAAATGA---AGGATGTAATGGCGGCTCATGGA 720
Qy 624 CCAAGCCTTTGAATACGTTAAAGTACAAATGAGGGCATCGACACAGACAGACTTATCCATA 683
Db 721 TTATGCTTTGAATTCATATCAACAACGAGGCAATGATACCGAGGAGATATATCCCTA 780
Qy 684 CTTGGTGTCAATGTATCTGCAACTTCAAGCAGAGAAATGTTGGTGTCAAGGTCAATGA 743
Db 781 TAGAGCTGTAGATAGCACTTGTGACCAATACAGGAAGAAACGCAAGGTTGTGACGATGA 840
Qy 744 TTCGATAACATCACCTCGGTGGTGGTGGATGAGTGGAGCATGCAAGTGGGCTTGGTGG 803
Db 841 CGATTTATGAGATGTTCCAGAAATGATGAGAAAGCATGTCAAAAGGCGATTTGCTAATCA 900
Qy 804 TCCAGTTAGCTGTGCAATTTGAGGTTTGAAGGTTTCAATCTGTACAGAAAGGTTGATA 863
Db 901 ACCAGTCACTGTGGCAATTTGAAGCAGGAGCGCGGAAATTCAGTATTTATGATTCGGGTAT 960
Qy 864 CAGCAGTGACCTCTGGAAGAGATCCAATGATGTGACACGACGAGTTCTTGGCGTGG 923
Db 961 ATTTACTGGCAATGTGGACAGCTC-----TGGATCATGGGTTTACTGCGAGTGG 1011
Qy 924 TTATGAGTTCGAGGAGCGGATTCCTTATTTGGCTCATCAAGAACTCATGGGTTACAAATG 983
Db 1012 ATATGCGACAGAAACGGAGTTGATTCTGGATAGTGAAGAACTCATGGGCGGTAGCTG 1071
Qy 984 GGGTGCAATGCTGCTACTTTAAGATGGAAC 1012
Db 1072 GGGAGAGCAAGGGTACATCAAAATGGCAC 1100

RESULT 5
US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-64

Query Match 15.0%; Score 160.2; DB 3; Length 1474;
Best Local Similarity 56.5%; Pred. No. 2.3e-39;
Matches 344; Conservative 0; Mismatches 253; Indels 12; Gaps 2;
Qy 402 CGATGCGGTGCTTCTGTAAACGAGGATTTGGAGGAGCAAGGATAGTGAGCCCTGTAA 461
Db 455 CGAGCTGGAGGTTCCGAGACCGTTGATTGGAGCGGAAGAGGCGGTTAATGCCATTAA 514
Qy 462 GGAACCAAGGAGCTGTGGATCTTGTGGACTTTTCACTACTTGGAGCTAGAGGCTGC 521
Db 515 AAACCAAGGAGCTTGGGAAAGTTTGGCGCTTTTCGACAGCTGCGAGCAGTAGAAGGTAT 574
Qy 522 ATATACAGCTAATCTGGAAGAGCAATCATTTATCTGACCAACTTTGTGACTGTGC 581

575 AACAAAGATCGTAAACAGGAGAACTCATATCTCTGTCGGAACAGAACTTGTGCACTGCGA 634
582 CTCAGCATCAATTAACCTTGGATGCAATGGAAGTTCCTTCCCAAGCCTTTGAATACGT 641
635 C---AGATCCTCAACCAAGGCTGCAACGGTGGTTTAAATGAGACTATGCTTTTCAATTCAT 691
642 TAAGTACAAATGGAGGATCGACACAGAAACAGACTTATCCATACCTTGGTGTCAATGGTAT 701
692 CATGAAGAAACGGCGGTTTGAACACCGAGCAAGATTAATCTTACCGTGGTTCCAAATGGA 751
702 CTCAACATCTCAAGCAGGAGAAATGTTGGTGTCAAGGTCAATGATTCGATAAACATCACCT 761
752 ATSCAATCTTACTCAAGAAATTCAGAGTGTGAAGTGTGAAGTGTGTTACGAAGATGTTCC 811
762 GGTGCTGAGGATGAGTGAAGCATGAGTGGCTTGGTGTGCTTCAAGTGTAGCTTGCAT 821
812 TACTGAAGATGAACCGGCTTGAAGAGAGCAGTTTCATACCGCCGCTGAGTGTGGCCAT 871
822 TGAGGTTGTGAAGGTTTCAATCTGTACAAGAAAGTGTATACAGCAGTGACACCTGTGG 881
872 TGAAGCTGGTGAAGAGTTCCTCAACATTACCAATCGGGATCTTCACTGGAAGAGTGTG 931
882 AAGAGATCCAAATGGATGTGAACACGAGTTCCTTGGCTGCTGCTTATGGAGTCAAGACGG 941
932 GACAAATC-----TAGATCATGAGTGGTGGCTGTGGTGTATGTTTCAGAGAACGG 982
942 GATTCTTATTGGCTCATCAAGAACTCATGGGCTACAAATTTGGGGTGACAAATGGCTACT 1001
983 TATTGACTATTGATTGTAAGGAATCTGGGGGTACAGTTGGGGAGAGGTGGTTACAT 1042
1002 TAAGATGGA 1010
1043 TAGGATGGA 1051

RESULT 6

US-09-120-365-6
; Sequence 6, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; EARLIER FILING DATE: 1998-07-22
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Periplaneta americana
US-09-120-365-6

Query Match 14.8%; Score 159; DB 3; Length 1833;
Best Local Similarity 51.0%; Pred. No. 6.1e-39;
Matches 462; Conservative 0; Mismatches 425; Indels 18; Gaps 3;
QY 166 TTCACAGGTTGCGCCGAGGTACGGGAAGAGCTACGGATCGAGGAGGAGATCAAGAAG 225
DB 812 TTTGATCATTCGAAAAAGACACAGCAAGACTACGCCAGCAATTTGGAACACACAAAG 871
QY 226 AGTTGCGGATCTTCGTTGGAATCTAGCGTTTATCCGGTCCACTAATCGGAAGGATCTG 285
DB 872 AGAAAAAGAAATTTTCCGACAGAAATTTGAGGTTTCATCCATTTAGAAATCGTGTAGATTA 931
QY 286 TCGTATACCTTAGGAATCAACCAATTCGCGACCTGACCTCGGGAGGAATTC----- 337
DB 932 GGATTCATCTGGAGCTGAACCAATTTGCGGACCGGACAGAGCTCGAACTGAAAGCTCTG 991
QY 338 ----GGACCAATCGCTTGTGGCGGCGAGAACTGTCTCGGGACTCGGATGGAACCCAC 393

DB 992 AGGGGAAGCAATACATCTGATGGGTACAATGGAGTTCTCCATTTCTTTATACCAATCTT 1051
QY 394 CGGTTTTCGATGGCTGCTTCTGTHAACGAGGATGGAGGAGCAAGGATAGTGAGC 453
DB 1052 GAGCAATCATGGACCAAAATTCCTGATGATTTGCTGAGGAGAAATTTATGTTGCTGACT 1111
QY 454 CCTGTAAGGACCAAGGAAGCTGTGGACTTCTGTGGAATTTCTAGTACTACTGGAGCACTA 513
DB 1112 CCAGTTAAGATCAGTCTGTTGGGCTCTTGTGGAGTTTGGGACTACTGGCACCATC 1171
QY 514 GAGCTGCATATACACAGCTAACTGGAAAGAGACATCATTTATCTGAACAGCACTTGTG 573
DB 1172 GAAGGAGCTTATTTCTTAAAGTATGGACATTTGGTGGATTTGTCAACACAGGCTCTAAT 1231
QY 574 GACTGTGCTCAGCATCAATAACTTTGGATGCAATGGAGGTTTCCCTTCCCAAGCCTTT 633
DB 1232 GACTGTGAGCTGGGCTATGTTAACAATGGTGTGATGGAAGTGAAGATTTCCGCTCTAT 1291
QY 634 GAATACCTTAAGTACAATGAGGATCGACACAGAAACAGACTTATCCA---TACCTTGGT 690
DB 1292 GAATGGATGATGAAGCATGGTGGCATCCCACTGGAAGACGAATATGGAGGCTATTTGGGC 1351
QY 691 GTCAATGCTATCTCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCAATGATTCGATA 750
DB 1352 CAGGATGCTATTTGCTGTCGAAAATGTACTCTCACAGCAAGATCACTGGCTATGTG 1411
QY 751 AACATCACTCCCTGGTGTGAGGATGAGTTGAAGCATGAGTGGCTTGTGGCTGCTCAAGTT 810
DB 1412 AACGTCATCTGAGACATTTGATGCTTTGAAGTAGCACTAGCAAGAGCATGGTCCCATC 1471
QY 811 AGCGTTGCATTTGAGTGTG---AAAGGTTTCAATCTGTACAGAAAGGTTGATACAGC 867
DB 1472 TCTGTGCAATTTGATGCTTCTCAAGACCTTCTCTTCTTATTCGAATGGCATCTACTAT 1531
QY 868 AGTGACACCTGTGAAGAGATCCAAATGGATGTGTAACCAAGCAGTCTTTCGCGTGGTAT 927
DB 1532 GATCTGAGTGTGGAACAAGCTAGATCAGTTGGACCAAGCAGTCTTCTGCTGGTTAT 1591
QY 928 GGAGTCGAGGAGGATTCCTTATTTGGCTCATCAAGACTCATGGGTACAAATGGGGT 987
DB 1592 GGCATCAATAATGAAACCCCTTATTTGGTGGTCAAAAATTCCTGCTTAATTTATTTGGGC 1651
QY 988 GACAATGGCTACTTTAAGATGGAATCGGCAAGCACTGTGGTGTGTTGCAACTTGGCA 1047
DB 1652 AACGATGCTACATCTTATGTCACCTAAGGATTAACAATGCTGTGTGTTGCAACTGACCT 1711
QY 1048 TCTTA 1052
DB 1712 ACCTA 1716

RESULT 7

US-09-515-039-6
; Sequence 6, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Periplaneta americana
US-09-515-039-6

Query Match 14.8%; Score 159; DB 3; Length 1833;
Best Local Similarity 51.0%; Pred. No. 6.1e-39;

Matches 462; Conservative 0; Mismatches 425; Indels 18; Gaps 3;
QY 166 TTCCGAGGTTCCGCCGAGGTACCGGAAGAGTACGATCGGAGGAGGATCAAGAG 225
Db 812 TTTGATCAGTTCCGAAAAGACACAGCAAGAGTACGCCAGCAATTTGGAAACACAAAG 871
QY 226 AGGTTCCGGATCTTCGTGAGAAATCTAGCGTTTATCCGTCCTCAATATCGGAAGAGTCTG 285
Db 872 AGAAAGAAATTTCCGACAGAAATTTGAGGTTTATCCATTTCAAGAAATCGTGTAGATTA 931
QY 286 TCGTATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGGAAATTC----- 337
Db 932 GGATTCATCTCTGGACGTGAACCAATTTGCCGACCGGACAGAGCTCGAAAGTCTGTG 991
QY 338 ----GGACCAATCCCTTGGTGGCGCGCAGAACTGCTCGCGACTGCGCATGGAACCCAC 393
Db 992 AGGGGGAAGCAATACATGATGGGTACATGGAGTTCTCCATTTCTTATCAATCTT 1051
QY 394 CGGTTTTCGATCGGCTTCCTGTAAACGAGGAAATGGAGGGAAGCAAGGATAGTGAAC 453
Db 1052 GACGCAATCATGGACCAAAATTCCTGATGATTTGCACTGGAGAAATTTATGGTGTGACT 1111
QY 454 CCTGTAAGGACCAAGGAGCTGTGGATCTTCTGGACTTTCAGTACTACTGGAGCACTA 513
Db 1112 CCAGTTAAGATCAGTCTGTTTGGCTCTTGTGGAGTTTGGGACTACTGGACCAATC 1171
QY 514 GAGGCTGCATATACACAGCTAACTGGAAGAGCAATCATTTATCTGAACAGCAACTTTGTG 573
Db 1172 GAAGGAGCTTATTTCTTAAAGTATGGACATTTGGTGGATTTGTCAACAGAGCTTAAT 1231
QY 574 GACTGTGCTCAGCATTCATTAATCTTGGATGATGAGGTTTCCCTCCCAAGCCTTT 633
Db 1232 GACTGTGAGCTGGGCTATGTTAAATGTTGATGGAGTGAAGATTTCCGCTCTTAT 1291
QY 634 GAATACGTTAAGTACAACTGAGGAGCATCGACACAGACAGACTTATCCA---TACCTTGGT 690
Db 1292 GAATGATGATGAAGCATGGTGGCATCCACTGGAAGACGAATATGGAGCTATTTGGGC 1351
QY 691 GTCAATGATATCTGCAACTTCAAGCAGGAGATGTTGGTGTCAAGGTCAATGATCGATA 750
Db 1352 CAGGATGGCTATTGTCTATGTGCAAAATGTAACTCTCAACGAAAGATCACTGGCTATGTG 1411
QY 751 AACATCACTCCGTGGTGTGAGATGAGTTGAAGCATGAGTGGGCTTGTGGTCCAGTT 810
Db 1412 AACGTCACTCGAGACATTTGATGCTTTGAGGTAGCATAGCAAGCATGTGTCCTATC 1471
QY 811 AGCGTTGCAATTTAGGTTGTG---AAAGGTTTCAATCTGTACAGAAAGGTGTATACAGC 867
Db 1472 TCTGTTGCAATGATGCTTCTCAAGACCTTCTCCTTCTATTCGAATGGCATCTACTAT 1531
QY 868 AGTGACACCTTGGAGAGATCCATGATGTGAACCAAGCAGTTCTTGGCGTGGTTAT 927
Db 1532 GATCCTGAGTGTGAAAACAGCTAGATCAGTTGACCAAGCAGGATCTGTGGTTGGTTAT 1591
QY 928 GGAGTTCGAGGACGGGATTCCTTATTTGGGTCTATCAAGAACTCATGGGTACAAATTTGGGT 987
Db 1592 GGCATCAATAATGAAACCTTATTTGGTGTGTCAAATTTCTGTCTAATTTATTTGGGC 1651
QY 988 GACAAATGCTATTTAAGATGGAATCTGGCAGAAATGATGTGTGTGTGCAACTTGGCA 1047
Db 1652 AACGATGGCTACATCTTATGTCACCTAAGGATAACCACTGTGTGTGTGCAACTGACCT 1711
QY 1048 TCTTA 1052
Db 1712 ACCTA 1716

RESULT 8

US-09-500-651-1
; Sequence 1, Application US/09500651
; GENERAL INFORMATION:
; APPLICANT: ASANO, MINAO
; APPLICANT: KAWAI, MISAOK

APPLICANT: MIWA, TETSUYA
APPLICANT: NIO, NORIKI
TITLE OF INVENTION: AMINOPEPTIDASE GK, AND A METHOD OF
TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSES: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSER: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/500,651
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/813,591
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 030458/1997
FILING DATE: 14-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-845-0
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1056 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1056
OTHER INFORMATION: /note= "METHOD OF DETERMINING THE
OTHER INFORMATION: CHARACTERISTICS: P"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..9-
OTHER INFORMATION: /note= "INSERTION SEQUENCE, METHOD
OTHER INFORMATION: OF DETERMINING THE CHARACTERISTICS: E"
US-09-500-651-1

Query Match 14.7%; Score 157.6; DB 1; Length 1056;

Best Local Similarity 52.7%; Pred. No. 1.2e-38;

Matches 455; Conservative 0; Mismatches 379; Indels 30; Gaps 4;

QY 166 TTCGACAGGTTCCGCCGAGGTACCGGAAGAGTACGATCGGAGGAGGATCAAGAG 225

Db 73 TAGGAGCAGTGGCTCGTGAAGCACGGAAGGTGTCAACGCGCTCGGCGGAGGAGAG 132

QY 226 AGGTTCCGGATCTTCGTGGAGAAATCTAGCGTTTATCCGGTCCACTAA---TCGGAGGAT 282

Db 133 CGCTTCCAGATCTTCAAGGACACACCTCGGATTCAGCAGCACCACAACTCCCGGAGGAC 192

QY 283 CTGTGCTATACCTTAGGAATCAACCAATTCGCGGACCTGACCTGGAGGAAATTCGGACC 342

Db 193 CGAACCTCAAGCTCGGACTGAACCGGTTGCTGTATCTCACCACGAGGAATACAGGGCC 252

QY 343 AATCGCCTGTTGTGGCGGCAGAA-----CTGCTCGCGGACTGCGCATGGA 387

Db 253 AAGTACTTGGGAACCAAGATGATCCCAACCGGAGGCTCGGAAGACACCCGAGCAACCGC 312
Qy 388 AACCAACCGGTTTGTGGTGGCTGCTTCTGTAAACGAGGATTTGGAGGACCAAGGATA 447
Db 313 TAGCGCCACGCTGTGGGACAAATTTGCTGATTCGGTTGATTGGAGGAAGAGGTGCT 372
Qy 448 GTGAGCCCTGTAAAGACCAAGGAGCTGTGGATCTTCTGCTGAGTTCAGTACTACTGGA 507
Db 373 GTTCTCTCTGTCAAAAGCAAGAGGCTGTGGAGCTGTGGCAATTTCTAGCAATCGT 432
Qy 508 GCATAGAGGCTGCATATACACAGCTAACTGAAAGAGCACATCATTTATCTGAACAGCAA 567
Db 433 GCAGTAGAGGATTAATAAGATAGTAACAGCGCACTGATTTCTGTTATCAGAACAGAA 492
Qy 568 CTTGTGAGCTGTCCCTCAGCACTTAATACTTTGATGCAATGGAGTTTSCCTTCCAA 627
Db 493 TTGTTGGATTGTGATACTGGAT---ATAACCAAGATGCAATGGAGGACTTATGGACTAT 549
Qy 628 GCTTTTGAATACGTTAAGTACAAATGAGGACATCGACACAGACAGACTTATCCATACCTT 687
Db 550 GAATTTGATTTATATCACAATGGCGCATTTGATGATGAGGATTAACCATACCGT 609
Qy 688 GGTGTCATGCTATCTGCACTTCAAGCAGGAGATGTTGGTGTCAAGGTCAATGATTCG 747
Db 610 GGTGTTGATGTTAGTGCACACATATAGGAAATGCTAAAGTCTGTTCTATTGATGAC 669
Qy 748 ATAACATCACCTGGTCTCAGGATGATTTGAAGCATGCACTGGGCTTGGTGGTCCA 807
Db 670 TAGGAAGATGTTCTGCTATATGATGATGATGATGATGATGATGATGATGATGATG 729
Qy 808 GTTAGCGTTGCAATTTGAGGTTGTGAAGGTTTCAATCTGTACAAAGAGGTGTATACAGC 867
Db 730 GTGAGCGTTGCTATTGAAGGAGGCGCAGGAATTTCAATATATGATGATGATGATG 789
Qy 868 AGTGACACCTGTGGAAGATCAATGATGATGATGATGATGATGATGATGATGATGATG 927
Db 790 ACGGGAGATGTGGCAGACAC-----TAGATCATGTTGTGTTGCTGTTGTTGTTAT 840
Qy 928 GGAGTCGAGACGGGATTCCTTATTTGGCTCATCAAGAACTCATGGGTACAAATTTGGGT 987
Db 841 GGAACAGCTAAAGTTCATGATTTGATGATGATGATGATGATGATGATGATGATGATG 900
Qy 988 GACAATGGCTACTTTAAGATGGA 1011
Db 901 GAGGATGGCTACATCAGATTAGAA 924

RESULT 9

US-08-813-591-1
; Sequence 1, Application US/08813591
; Patent No. 5824534
; GENERAL INFORMATION:
; APPLICANT: ASANO, MINAO
; APPLICANT: KAWAI, MISAKO
; APPLICANT: MIWA, TETSUYA
; APPLICANT: NIO, NORIKI
; TITLE OF INVENTION: AMINOPEPTIDASE GX, AND A METHOD OF
; TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P. C.
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,591
FILING DATE: 07-MAR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 051848/1996
FILING DATE: 08-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 030458/1997
FILING DATE: 14-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-845-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1056 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1056
OTHER INFORMATION: /note= "METHOD OF DETERMINING THE
OTHER INFORMATION: CHARACTERISTICS: P"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..9
OTHER INFORMATION: /note= "INSERTION SEQUENCE, METHOD
OTHER INFORMATION: OF DETERMINING THE CHARACTERISTICS: E"
US-08-813-591-1

Query Match 14.7%; Score 157.6; DB 1; Length 1056;
Best Local Similarity 52.7%; Pred. No. 1.2e-38;
Matches 455; Conservative 0; Mismatches 379; Indels 30; Gaps 4;

Qy 166 TTCACAGCGTTTCGCCCGCAGGTACGGGAAGAGCTACGGATCGGAGGAGGAGATCAAGAAG 225
Db 73 TACGAGCAGTGGCTCGTGAAGCAGCGGAGGTGTACAAACGCTCGCGGAGAGAGAG 132
Qy 226 AGTTTCGGGATCTTCGTGGAGAACTAGCGTTTATCCGTTCCACTAA---TCGGAAGAT 282
Db 133 CGCTTCCAGATCTTCAAGGACCACTCGGATTCATCGACCACTCCGCGGAGGAC 192
Qy 283 CTGTCTATACCTTAGGAATCAACCAATTCGCGACCTGACCTGGGAGGAAATTCGGAC 342
Db 193 CGAACCTACAGCTCGGACTGACCGGTTCCGCTGATCTCACCACGAGGAATACAGGGCC 252
Qy 343 AATCGCTTGTGGCGGCGAGAA-----CTGCTCGGCGACTGCGCATGGA 387
Db 253 AAGTACTTGGGAACCAAGATCGATCCCAACCGGAGGCTCGGAAGAGACCCGAGCAACCGC 312
Qy 388 AACCAACCGGTTTGTGATGCGTCTTCTGTAAACGAGGATTTGGAGGACCAAGGATA 447
Db 313 TAGCGCCACGCTGTGGGACAAATTTGCTGATTCGGTTGATTGGAGGAAGAGGTGCT 372
Qy 448 GTGAGCCCTGTAAAGACCAAGGAGCTGTGGATCTTCTGCTGAGTTCAGTACTACTGGA 507
Db 373 GTTCTCTCTGTCAAAAGACCAAGGAGGCTGTGGAGCTGTTGGGCAATTTCTAGCAATCGT 432
Qy 508 GCATAGAGGCTGCATATACACAGCTAACTGAAAGAGCACATCATTTATCTGAACAGCAA 567
Db 433 GCAGTAGAGGATTAATAAGATAGTAACAGCGCACTGATTTCTGTTATCAGAACAGAA 492
Qy 568 CTTGTGAGCTGTCCCTCAGCACTTAATACTTTGATGCAATGAGGTTTSCCTTCCAA 627
Db 493 TTGTTGGATTGTGATACTGGAT---ATAACCAAGATGCAATGGAGGACTTATGGACTAT 549

628	GCCTTTGAATACGTTTAAGTACAATAGGAGGCATGCACACAGAACGACTTATTCATACCTT	687
	Db	
550	GCAATTTGAGTTTCATAATCAACAATAGCGGCATTTGATTTCTGATGAGGATTACCCATACCGT	609
	Qy	
688	GGTGTCAAATGGTATCTCGAACTTCAAGCAGGAGAAATGTGGTGTCAAGGTCAATTTGATTGG	747
	Db	
610	GGTGTGTGATGGTAGATGCGCACATATAGGAAAATGCTAAAGTCGTTCTTATTGATGAC	669
	Qy	
748	ATAAACATCACCCCTGGGTCTGAGGATGAGTTGAAGCATGCAGTGGGCTTTGGTGCCTCCA	807
	Db	
670	TACGAAGATGTTCTCGCTATGATGAGTTAGCCCTTGAAAAGGCCGTTGCAAAATCAGCCC	729
	Qy	
808	GTTAGCGTTGCATTTGAGGTTGTGAAGTTCCTAACTCTGTACAAGAAAGGTGTATACAGC	867
	Db	
730	GTGAGCGTTGCTATTGAAAGAGGGGGCAGGGAAATTCAAATTATATGTATCTGGTGTATTCC	789
	Qy	
868	ACTGACACCTGTGGAAAGATCCCAATGGATGTGAACCAACGACAGTCTTTGCCGTCGGTTAT	927
	Db	
790	ACGGGAGATGTGGCACAGCAC-----TAGATCATGGTGTGCGCTGTTGGGTAT	840
	Qy	
928	GGAGTCGAGGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAAATGGGGGT	987
	Db	
841	GGAAAGCTTAAGGTCATGATTTATGGATCTAGAGTAATCATGGGGTCTTACGTCGGGA	900
	Qy	
988	GACAAATGGCTACTTTAAGATGGAA	1011
	Db	
901	GAGGATGGCTACATCAGATTAGAA	924

```

RESULT 10
US-09-325-932A-112
; Sequence 112, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE OF INVENTION: death and their use in the modification of forestry plant develop
; FILE REFERENCES: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
; US-09-325-932A-112

```

Query Match	14.5%;	Score 155;	DB 4;	Length 522;
Best Local Similarity	63.8%;	Pred. No. 5.1e-38;		
Matches 252;	Conservative 0;	Mismatches 140;	Indels 3;	Gaps 1;
Qy	116	TTGAGCTGTGCATCCTCGGGCTCTCTGGCAGTTGCCGCCACGCCCTTCCACTTCGCACGGT	175	
Db	128	TTGAGTCTCCATCTCTCAACCGTGGCCACACCGTCCGGCCCTCTCTCTTGGTAGACT	187	
Qy	176	TCGCCCCGAGGTACGGGAAGAGCTACGGATCGGAGGAGGAGATCAAGAAGAGGTTTCGGA	235	
Db	188	TCGCCCCGCGGCACGGGAAGACTTACAAGACAGCGGAGGAGATCAAGTTGAGGTTTCGATA	247	
Qy	236	TCCTTCGTGGAGAACTAGCGGTTTATCCGGTCCACTAATCGGAAGGATCTGTGCTATACCC	295	
Db	248	ATTACAGGAGAACTCAAGTCTCATTCGATCCACCAACGAAGGGCTTGCTTTACACCTC	307	
Qy	296	TAGGAATCAACCAATTCCGCGACTGACCTGGGAGGAATTCGGACCAATCGCCTTCGGT	355	
Db	308	TCGCTGTTAATCAGTATGCTGACTGGAGCTGGGAGGATTCAGACGCCACAGACTGGAG	367	
Qy	356	CGGCGCAGAACTGCTCGCGGACTCCGCATCGAAACCAACCGGTTTGTTCAGTGGGCGTTC	415	
Db	368	CTTCTCAAGACTGTCTTGCCACCAACCAAGGCGACCAAGCTCACCGAGATGTTCTTC	427	

QY 416 CTCTTAACAGAGGATTCGAGGAGCAAGCGATAGTGAG---CCCTGTAAAGGACCAAGGAA 477
Db 428 CCGAACAACGAAGACTGGGAGAGAAAGGGCAATTGTAGCCCCCAGTTAATAGATCAAGGCGG 487
QY 473 GCTGTGGATCTTGTCTGGAGCTTTTCAGTACTACTGGA 507
Db 488 CTGTGGATCTTGTCTGGAGTTTCAGCGCAACTGGA 522

RESULT 11
US-08-883-526-2
; Sequence 2, Application US/080883526
; Patent No. 6033893
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Guebler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW HUMAN CATHEPSIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,526
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0331 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1366 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: THYMOT02
; CLONE: 347021
; US-08-883-526-2

	Query Match	14.1%	Score 150.6;	DB 3;	Length 1366;
	Best Local Similarity	51.9%;	Pred. No. 2e-36;		
	Matches 392;	Conservative 0;	Mismatches 354;	Indels 9;	Gaps 2;
Qy	279	GGATCTGTCGTATACCTTAGGAATCAACCAATTTCGGCGACCTGACCTGGGAGGAATTCG	338		
Db	266	GAACCTTGGCTTCACAATGGCCATGAATGTTTTGGTGACATCAACCAATGAAGAATTCAG	325		
Qy	339	GACCAATCGGCCTTGGTGGCGCGCAGAACTCTCTGGGCGACTCGGCATGGAAACCAACCGGTT	398		
Db	326	GCAGATCATGGGTGGCTTTTCGAACCAACGAAATTCAGGAAGGGGAAAGTGTTCGTGTGAGCC	385		
Qy	399	TGTCGATGGCGCTCCTCTCTATACGAGGGAATTCGAGGGGACCAAGGATACGTAGCCCTGT	458		

386 TCTGTTTCTTGTATCTTCCAAATCTGTGGATTGGAGAAAGAGGCTACGTGACGCCAGT 445
459 AAAGACCAAGAGCTGTGGATCTTCTGTGACTTTCAGTACTACTGAGCACTAGAGGC 518
446 GAAGAATCAAGAACAGTGTGGTCTTCTGTGGGCTTTAGTGGCAGTGTGTCTTGAAGG 505
519 TGCATATACACAGCTAACTGGAAGAGACACATCTATCTGAACAGCAACTTGTGGACTG 578
506 ACAGATGTTCCGGAACACTGGGAACCTGTCTCACTGAGCGAGCAGATCTGGTGGACTG 565
579 TGCTCAGCATCAATAACTTGGATGCAATGGAGGTTTGCCTTCCCAAGCCCTTGAATA 638
566 TTGCGCTCTCAAGCAATCAGGCTGCAATGGTGGCTTCATGGCTAGGCGCTTCCAGTA 625
639 CGTTAAGTCAATGAGGATCGACACAGCAAGACTTATCCATACCTTGGTGTCAATGG 698
626 TGTCAAGGAGACGAGGCTGCACTCTGAGGAATCTTATCCATATGATGAGTGGATGA 685
699 TATCTCAACTTCAAGCAGAGAAATGTTGGTGTCAAGGTCAATGATTCGATAAACATCAC 758
686 AATCTGTAAGTACAGACCTGAGAAATCTGTCTAATGACACTGGGCTTCACAATGGTCGC 745
759 CTTGGGTGTGAGTGTGAGTGTGAAGCATGCTGGGCTTGGTGGCTTCCAGTTAGGCTTGC 818
746 ACTTGAAGGAGAGAGGCTGATGAAGCAGTGTGCAACTGTGGGCTTCCAGTGTGC 805
819 ATTGAGGTGTGAA---AGGTTTCAATCTGTACAGAAAGGTGTATACAGCAGTGACAC 875
806 TATGATGACAGGCAATTCGCTTCCAGTCTTACAAATCAGGCAATTTATTTGAACACAGA 865
876 CTGTGAAGAGATCCAAATGATGTGAACCAAGGCTTCTTGGCTGGTGTATGAGTGA 935
866 CTGCAGCAGCAAAACCTGGATCANGGTGTCTGTGGTGTGGCTACGGCTTTGAAGGAGC 925
936 GGACGGGATTCCT-----TATTGGCTCATCAAGAACTCATGGGTCACAAATGGGGTGA 989
926 AATTCGAATACAGCAAGTATTGGCTCTGCAAAACAGCTGGGGTCCAGATGGGGTGC 985
990 CAATGGCTATTGAAGTGAAGTGGGCTGGGCAAGCA 1024
986 GAATGGCTATTGAAGTGAAGTGGGCTGGGCAAGCA 1020

RESULT 12

US-09-519-283-2

Sequence 2, Application US/09519283

Patent No. 6620606

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Guegler, Karl J.

Corley, Neil C.

Shah, Purvi

TITLE OF INVENTION: NEW HUMAN CATHEPSIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/519,283

FILING DATE: 07-Mar-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/883,526

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PP-0331 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1366 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: THYMOT02

CLONE: 347021

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-519-283-2

Query Match

Best Local Similarity 14.1%; Score 150.6; DB 4; Length 1366;

Matches 392; Conservative 0; Mismatches 354; Indels 9; Gaps 2;

Qy 279 GGATCTCTCTATACCTAGGAATCACCAATTCGCGACCTGACCTGGGAGCAATTCG 338
Db 266 GAAACTTGGCTTCAATGGCCATGATGCTTTTGTGACATGACCAATGAAGATTGAG 325
Qy 339 GACCAATCGCTTGGTGGCGGCGAGAACTCTCTCGGACCTGCGATGGAACACCGGTT 398
Db 326 GCAGATGATGGTGTCTTCGAAACCAAGAAATTCAGGAAGGGAAGTGTTCGCTGAGCC 385
Qy 399 TGTGATGGGTGCTTCTCTGACGAGGATTCGAGGAGGAGGAGGAGGAGGAGGAGG 458
Db 386 TCTGTTTCTTGTATCTTCCAAATCTGFGATGAGAAAGAAAGGCTACGTGAGCCAGT 445
Qy 459 AAAGACCAAGGAGCTGTGGATCTTCTGCGACTTTTCAAGTACTACTGAGCAGCTAGAGG 518
Db 446 GAGAAATCAGAAACAGTGTGTTCTTCTGCGCTTTTGTGCGACTGCTGTCTTGAAGG 505
Qy 519 TGCATATACAGCTAATCGAAGAGCAGACATCTATCTGAAACAGCAACTGTGGAAGT 578
Db 506 ACAGATGTTCCGGAACACTGGGAAACTTGTCTCACTGAGGAGCAGAAATCTGTGGAGT 565
Qy 579 TGCCTCAGCATTCATTAATCTTGGATCAATGGAGGTTTGGCTTCCCAAGCCTTTGAATA 638
Db 566 TTGCGCTCTCAAGCAGCAATCAGGCTGCAATGGTGGCTTTCATGGCTAGGCGCTTCCAGTA 625
Qy 639 CGTTAAGTCAATGGAGGAGTTCGACACAGAAACAGACTTATCCATACCTTGGTGTCAATGG 698
Db 626 TGTCAAGGAGAAACGAGGCGCTGGAATCTGAGGAATCTCTATCCATATGACAGTGGATGA 685
Qy 699 TATCTGCAACTTCAAGCAGGAGATGTTGGTGTCAAGGTCAATGATTCGATTAACATCAC 758
Db 686 AATCTGTAAGTACAGACTGAGAAATCTGTGCTAATGACACTGGCTTCAATGGTGGCTGC 745
Qy 759 CTTGGGTGCTGAGGATGAGTTGAAGCAGTGGGCTTGGTGGCTGCTGAGTGGCTTGGCT 818
Db 746 ACCTGGAAGGAGAAAGGCGCTGATGAAAGCAGTGCCTGCAACTGTGGGCGCCATCTCCGTTGC 805
Qy 819 ATTTGAGGTGTGAA---AGGTTTCAATCTGTCAAGAAAGGTGTATACAGCAGTGACAC 875
Db 806 TATGATGAGGCGCAATTCGCTTCCAGTCTTCAAAATCAGGCAATTTATTTGAACACAGA 865
Qy 876 CTGTGGAAGAGATCCAAATGATGTGAACACGAGTCTTTCGCCCTCGGTTATGAGTGA 935
Db 866 CTGCAGCAGCAAAACCTGGATCATGGTGTCTGCTGGTGGCTTGGCTACGGCTTTGAAGGAGC 925
Qy 936 GGACGGGATTCCT-----TATTGGCTCATCAAGAACTCATGGGTCACAAATGGGGTGA 989
Db 926 AATTCGAATACAGCAAGTATTGGCTCTGCAAAACAGCTGGGGTCCAGATGGGGTGC 985
Qy 990 CAATGGCTATTGAAGTGAAGTGGGCTGGGCAAGCA 1024

Db 986 GAAATGGCTATGTAAAAATAGCCAAAGACAAAGAACA 1020
|||||

RESULT 13

US-08-821-994-61
; Sequence 61, Application US/08821994A
; Patent No. 6228643

GENERAL INFORMATION:

; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP

; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters

; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A

; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729

; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9

; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61

; LENGTH: 1390
; TYPE: DNA

; ORGANISM: Brassica napus
US-08-821-994-61

Query Match 13.8%; Score 147.4; DB 3; Length 1390;
Best Local Similarity 55.2%; Pred. No. 2e-35;

Matches 336; Conservative 0; Mismatches 261; Indels 12; Gaps 2;

QY 402 CGATGGCGTGCTCTCTGTAAACGAGGATTTGAGGGAGCAAGGGATAGTGAGCCCTGTAAA 461
|||||

Db 467 CGACGTGGAGTTCCGGAGACGGTTGACTGGAGACAGAGGAGCCGTTAATGCCATTAA 526
|||||

QY 462 GGACCAAGGAAGCTGTGGATCTTGTCTGGACTTTTCAGTACTACTGAGGACCTAGAGGCTGC 521
|||||

Db 527 AAACCAAGGACTTTGCGGAAGTTGTGGCGGTTTTCACACAGCTGCAGCAGTAGAAGGCAT 586
|||||

QY 522 ATATACACAGCTAACTGGAAGAGACACATATTCTGAAACAGCACTTTGGAAGCTGTC 581
|||||

Db 587 AAACAGATCGTAACAGAGAGCTCATATCTCTGTCGACAGAACTTTGCTGACTGCGA 646
|||||

QY 582 CTCAGACTTCAATTAATTTGGATGCAATGGAGGTTTGCCCTTCCCAAGCCTTTGAATAGT 641
|||||

Db 647 CAATCAT---ACAACCAAGGCTGTAAACGGCGGCTTAATGGATTATGCTTTCAATTAT 703
|||||

QY 642 TAAGTACAAATGGAGCATCGACACAGACAGACTTATCCATACCTTGGTGTCAATGGTAT 701
|||||

Db 704 CATGAAAACCGCGGATTTAAACACCGAGCAAGACTATCTCTTACCACGGAACCAATGGCAA 763
|||||

QY 702 CTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCATTTGATTCGATAAACATCACCT 761
|||||

Db 764 ATGCAACTCTTTACTTAAATTCGAGAGTTGTGACTATCGATGGATACGAAGATGTTCC 823
|||||

QY 762 GGGTCTGAGGATGAGTTGAAGCATGCAAGTGGGCTTGGTGGCTCCAGTTAGCGTTGCAAT 821
|||||

Db 824 TAGTAAAGATGAAACCGCGTTTGAAGAGAGCAGTTTCGTACCAGCCTGTGAGTGTGCTAT 883
|||||

QY 822 TGAGGTTGTGAAGGTTTCAATCTGTACAGAAAGGTTATACAGCAGTGCACACCTGTGG 881
|||||

Db 884 TGATGCTGGTGAAGAGCTTCCACATTACCATCTGGATCTTCTACTGGAAGTGTGG 943
|||||

QY 882 AAGAGATCCAATGGATGTAACACGAGTTCTTCCGCTCGGTTATGGATTCGAGGACGG 941
|||||

Db 944 -----TAGCACTATGGATCAGCTGTTGTGGCGGTTGGTTATGGATCAGAGAACGG 994
|||||

QY 942 GATTCCTTATTTGGCTCATCAAGAACTCATGGGGTACAAATTTGGGGTGCACATGGCTACTT 1001
|||||

Db 995 TGTGACTATTGGATTGTACGTAACTCTTGGGGTACAAAGCTGGGGAGAGGATGTTTACAT 1054
|||||

QY 1002 TAAGATGGA 1010

Db 1055 TAGGATGGA 1063
|||||

RESULT 14

US-08-821-994-63

; Sequence 63, Application US/08821994A
; Patent No. 6228643

GENERAL INFORMATION:

; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP

; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters

; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A

; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729

; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9

; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63

; LENGTH: 1441
; TYPE: DNA

; ORGANISM: Brassica napus
US-08-821-994-63

Query Match 13.8%; Score 147.4; DB 3; Length 1441;
Best Local Similarity 55.2%; Pred. No. 2.1e-35;

Matches 336; Conservative 0; Mismatches 261; Indels 12; Gaps 2;

QY 402 CGATGGCGTGCTCTCTGTAAACGAGGATTTGAGGGAGCAAGGGATAGTGAGCCCTGTAAA 461
|||||

Db 453 CGACGTGGAGTTCCGGAGACGGTTGACTGGAGACAGAGGAGCCGTTAATGCCATTAA 512
|||||

QY 462 GGACCAAGGAAGCTGTGGATCTTGTCTGGACTTTTCAGTACTACTGAGGACCTAGAGGCTGC 521
|||||

Db 513 AAACCAAGGACTTTGCGGAAGTTGTGGCGGTTTTCACACAGCTGCAGCAGTAGAAGGCAT 572
|||||

QY 522 ATATACACAGCTAACTGGAAGAGACACATATTCTGAAACAGCACTTTGGAAGCTGTC 581
|||||

Db 573 AAACAGATCGTAACAGAGAGCTCATATCTCTGTCGACAGAACTTTGCTGACTGCGA 632
|||||

QY 582 CTCAGACTTCAATTAATTTGGATGCAATGGAGGTTTGCCCTTCCCAAGCCTTTGAATAGT 641
|||||

Db 633 CAATCAT---ACAACCAAGGCTGTAAACGGCGGCTTAATGGATTATGCTTTTCAATTAT 689
|||||

QY 642 TAAGTACAAATGGAGCATCGACACAGACAGACTTATCCATACCTTGGTGTCAATGGTAT 701
|||||

Db 690 CATGAAAACCGCGGATTTAAACACCGAGCAAGACTATCTCTTACCACGGAACCAATGGCAA 749
|||||

QY 702 CTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCATTTGATTCGATAAACATCACCT 761
|||||

Db 750 ATGCAACTCTTTACTTAAATTCGAGAGTTGTGACTATCGATGGATACGAAGATGTTCC 809
|||||

QY 762 GGGTCTGAGGATGAGTTGAAGCATGCAAGTGGGCTTGGTGGCTCCAGTTAGCGTTGCAAT 821
|||||

Db 810 TAGTAAAGATGAAACCGCGTTTGAAGAGAGCAGTTTCGTACCAGCCTGTGAGTGTGCTAT 869
|||||

QY 822 TGAGGTTGTGAAGGTTTCAATCTGTACAGAAAGGTTATACAGCAGTGCACACCTGTGG 881
|||||

Db 870 TGATGCTGGTGAAGAGCTTCCACATTACCATCTGGATCTTCTACTGGAAGTGTGG 929
|||||

QY 882 AAGAGATCCAATGGATGTAACACGAGTTCTTCCGCTCGGTTATGGATTCGAGGACGG 941
|||||

Db 930 -----TAGCACTATGGATCAGCTGTTGTGGCGGTTGGTTATGGATCAGAGAACGG 980
|||||

QY 942 GATTCCTTATTTGGCTCATCAAGAACTCATGGGGTACAAATTTGGGGTGCACATGGCTACTT 1001
|||||

Db 981 TGTGACTATTGGATTGTACGTAACTCTTGGGGTACAAAGCTGGGGAGAGGATGTTTACAT 1040
|||||

QY 1002 TAAGATGGA 1010

Db 1041 TAGGATGA 1049

Db 878 TAGGATGA 886

Search completed: March 29, 2004, 21:10:30
Job time : 83 secs

RESULT 15
US-08-821-994-86
; Sequence 86, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jenson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
; LENGTH: 1102
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-86

Query Match 13.6%; Score 145.8; DB 3; Length 1102;
Best Local Similarity 55.0%; Pred. No. 5.5e-35;
Matches 335; Conservative 0; Mismatches 262; Indels 12; Gaps 2;
QY 402 CGATGCGCTGCTTCCTGTAACGAGGATGGAGGAGCAAGGATAGTGAGCCCTGTAAA 461
Db 290 CGACGTGGAGGTTCCGGTACCGTTGACTGGAGACAGAAAGAGCGGTTATGCCATTAA 349
QY 462 GGACCAAGGAAGCTGTGGATCTTGTGGACTTTTCAGTACTACTGGAGCACTAGAGGCTGC 521
Db 350 AGACCAAGGAACCTTGGGAAGTTTGGCGGTTTTCAACAGCTGCAGCAGTAGAAGGTAT 409
QY 522 ATATACACAGCTTAACGGAAGAGCACATCATCTGACAGCAACTTGTGGACTGTGC 581
Db 410 AAACAAGATCGTAACAGGAGAACTCGTATCTTTGTCGGAACAGAACTTGTGACTGGCA 469
QY 582 CTCAGCATTCATTAATCTTTGGATGCAATGGAGGTTTGGCCTTCCCAAGCCTTTGAATAGCT 641
Db 470 CAATCGT---ACAACCAAGGCTGTAAACCGGCTCTAATGGATTATGCTTTTCAATTCA 526
QY 642 TAAGTCAATGGAGGCAATCGACACAGACAGACTTATCCATACCTTGGTGTCAATGGTAT 701
Db 527 AATGAATAACGGCGGATTAAACACCGAGAAAGACTATCCTTACCAGGAAACCAATGGCAA 586
QY 702 CTGCAACTTCAAGCAGGAGAACTGTGGTGTCAAGGTCATTGATTCGATAAACATCACCT 761
Db 587 ATGCAACTCTTTACTTAAGATTCAGAGTTGTAACTATCGATGGATACGAGATGTTCC 646
QY 762 GGGTGTGAGGATGAGTGAAGCATGAGTGGGCTTGGTGGTCCAGTTAGCGTTGCAATT 821
Db 647 TAGTAAAGATGAACCGCGTTGAAGAGAGCAGTTTCATACAGCCTGTGAGTGTGCTAT 706
QY 822 TGAGGTTGTGAAGGTTTCAATCTGTACAGAAAGGTGTATACAGAGTGACACCTGTGG 881
Db 707 TGATGCTGTGGAAGAGCTTTTCCAACTTACCATTCTGGAATCTTCACTGGAAAGTGG 766
QY 882 AAGAGATCCAATGGATGTGAACCAACGAGTTCTTTCGCGTCCGTTATGGAGTCGAGGACGG 941
Db 767 -----TACGAATATGATCACGCTGTGGTGGCGGTTGGTTATGGGTACAGAACGG 817
QY 942 GATTCCTTATGGCTCATCAGAACTCATGGGTACAAATTTGGGTGACAAATGGCTACTT 1001
Db 818 CGTTGACTATTGAGTTGACGTAACCTTTGGGGTACACGTTGGGGAGGAGGATGGTTACAT 877
QY 1002 TAAGATGA 1010

Db	61	GC	CGGTTTCG	AAGACAATCCAAATCCGGTCCGTTACAAAGGCGCTGACTCGAATGAG	120
Qy	121	CCT	GCCATCCCTCGCGCTCCTTTGGCAGTTGCCGCCACGCTTCCACTTCGACCGGTTCCGC	180	
Db	121	CCT	GCCATCCCTCGCGCTCCTTTGGCAGTTGCCGCCACGCTTCCACTTCGACCGGTTCCGC	180	
Qy	181	CGC	AGGTACCGG	AGAGACTACGATCCGAGCGAGGAGAGATCAAGAGAGGTTCCGGGATCTTC	240
Db	181	CGC	AGGTACCGG	AAAGACTACGATCCGAGCGAGGAGATCAAGAGAGGTTCCGGGATCTTC	240
Qy	241	GTG	GAGAACTCTAGCGTTTATCCGGTCCACTAATCCGGAAGGATCTGTCTGATACCCCTAGGA	300	
Db	241	GTG	GAGAACTCTAGCGTTTATCCGGTCCACTAATCCGGAAGGATCTGTCTGATACCCCTAGGA	300	
Qy	301	ATCA	ACCAATTCGCGGACCTGACCTGGAGGAATTCGGGACCAATCGCTTGGTGGCGG	360	
Db	301	ATCA	ACCAATTCGCGGACCTGACCTGGAGGAATTCGGGACCAATCGCTTGGTGGCGG	360	
Qy	361	CAG	AATGCTCGGCGACTGGCGCATGGGAAACACCGGTTTTCGATGCGCTTCCTGTGA	420	
Db	361	CAG	AATGCTCGGCGACTGGCGCATGGGAAACACCGGTTTTCGATGCGCTTCCTGTGA	420	
Qy	421	ACG	AGGATTTGGAGGGAGCAGGATAGTGAGCCCTCTAAGGACCAAGGAGCTGGGA	480	
Db	421	ACG	AGGATTTGGAGGGAGCAGGATAGTGAGCCCTCTAAGGACCAAGGAGCTGGGA	480	
Qy	481	TC	TGCTGGACTTTTCAGTACTACTCTGGAGCACTAGAGGCTGCATATACACAGCTAACTGGA	540	
Db	481	TC	TGCTGGACTTTTCAGTACTACTCTGGAGCACTAGAGGCTGCATATACACAGCTAACTGGA	540	
Qy	541	AAG	AGCACATCATTTATCTGAACAGCAACTTGTGACCTGCGCTCAGCATTCAAATACTTT	600	
Db	541	AAG	AGCACATCATTTATCTGAACAGCAACTTGTGACCTGCGCTCAGCATTCAAATACTTT	600	
Qy	601	GGAT	GCAATGGAGGTTTTCGCTTCCCAAGCCCTTGAATACGTTAAGTACAATGGAGGCATC	660	
Db	601	GGAT	GCAATGGAGGTTTTCGCTTCCCAAGCCCTTGAATACGTTAAGTACAATGGAGGCATC	660	
Qy	661	GAC	CAGAAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAG	720	
Db	661	GAC	CAGAAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAG	720	
Qy	721	AAT	GTTGGTGTCAAGGTCATTGATTCGATAAACATCACCTTGGGTGTCTGAGGATGAGTTG	780	
Db	721	AA	TGTTGGTGTCAAGGTCATTGATTCGATAAACATCACCTTGGGTGTCTGAGGATGAGTTG	780	
Qy	781	AAG	CATCAGTGGGCTTTGGTGGCTGCTAGCGTTGCAATTTGAGGTTTGAAAGGTTTC	840	
Db	781	AAG	CATCAGTGGGCTTTGGTGGCTGCTAGCGTTGCAATTTGAGGTTTGAAAGGTTTC	840	
Qy	841	AAT	CTGTACAAGAAAGGTGTATACAGCAGTGACACCTGTGGAAGAGATCCAAATGGAATGTG	900	
Db	841	AA	TCGTACAAGAAAGGTGTATACAGCAGTGACACCTGTGGAAGAGATCCAAATGGAATGTG	900	
Qy	901	AA	CAACGAGTTCTTTGGCGGTATATGGAGTCAAGGACGGGATTCCTTATTTGGCTCATC	960	
Db	901	AA	CAACGAGTTCTTTGGCGGTATATGGAGTCAAGGACGGGATTCCTTATTTGGCTCATC	960	
Qy	961	AAG	AACTCATGGGGTACAAATTTGGGGTGAACAATGGCTACTTTAAGATGGAACTCGGCAAG	1020	
Db	961	AAG	AACTCATGGGGTACAAATTTGGGGTGAACAATGGCTACTTTAAGATGGAACTCGGCAAG	1020	
Qy	1021	AAC	ATGTGTGGTGTTCGAACTTGGCGCATCTTATCCCAATTTGGCTGTGTAG	1071	
Db	1021	AAC	ATGTGTGGTGTTCGAACTTGGCGCATCTTATCCCAATTTGGCTGTGTGTAG	1071	

RESULT 2

RESULT 2
US-10-425-114-156

US-10-425-114-136
: Sequence 156, Application US/10425114

Sequence 136, Application US/10
; Publication No. US20040034888A1

; FUNCTIONAL NO: 0320
; GENERAL INFORMATION:
; GENERAL INFORMATION:

APPLICANT: Liu, Jingdong

Db	805	TGTTGGAGTCAAGGTTTTGGACTCGGTTAACA	CACCCTGGGTCTGAGATGAAC	864
Qy	783	GCATGCAGTGGGCTTGTTGGCTCCAGTTAGCGTTGCAATTTGAGGTTGTGAAGGTTTCAA	842	
Db	865	GGATGCTGTTGGTCTGGTTGCCCAGTTTACGTTGGCTTC	CGAGTGA	924
Qy	843	TCCTGTACAAGAAAGGTCTATACAGCAGTGCACCTGTGGAGAGATCCAATGATGTGAA	902	
Db	925	GCTGTACAAGACGGAGTTTACACTCGGACCAATTTGTGGAACTACACCGATGGATGTGAA	984	
Qy	903	CCACGCAGTTCTTGGCCGCTTATGAGTTCGAGACGGGATTCCTTATTTGGCTCATCAA	962	
Db	985	CCACGCTGTTCTGGCTGTGGCTACGGTTCGAAGATGGTGTCCCTACTGGGCTCATCAA	1044	
Qy	963	GAACCTCATGGGTACAATTTGGGGTGCACATGCTCTTTAAGATGAACTCCGGCAGAA	1022	
Db	1045	GAACCTCATGGGGCGCTGACTGGGGCGATGAGGGTTACTTTCAGATGAAATGGCGAAGAA	1104	
Qy	1023	CATGTGPGGTGTTGCACACTTTCGCGCATCTTATCCCAATTGTGC	1064	
Db	1105	CATGTGCGGTTGTGCTACGTGTGCATCTCACTACCTATTGTGCG	1146	

RESULT 3

US-10-425-114-14419
; Sequence 14419, Application US/10425114
; Publication No. US20040034888A1

; FIDUCIARY NO: 09200400348888
; GENERAL INFORMATION:

; GENERAL INFORMATION:
: APPLICANT: Liu Jingsong

APPLICANT: LIU, JINGDONG
APPLICANT: ZHOU, YIYUAN

APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David

; APPLICANT: Screen, Steve

APPLICANT: Tabaska, Jac

; APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT

FILE REFERENCE: 38-21(632313)B

FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: USC/1

; CURRENT APPLIC

; CURRENT FILING

; NUMBER OF S

; SEQ ID NO 14419

; LENGTH:

TYPE: DNA

ORGANISM: Zea m

Query Match	53.2%;	Score 569.6;	DB 12;	Length 1413;
Best Local Similarity	72.4%;	Pred. No. 7.7e-181;		
Matches 769; Conservative	0;	Mismatches 284;	Indels 9;	Gaps 2;

Qy	12	GCTCTCTTCTTCTTACTTCTTCTGTGTCCTCGCCCTCTCGTCGCGCTCGCCGTTTCGA	71
Db	93	GTCGTCTCTCGCGTCGTGCGCTTCGCGGCACGCGCGCGGCACATCCGCGTTTCGC	152
Qy	72	AGAAGACAATCCAAATCCGGTCCGTATACAAAGGCTTGACT---CGATTGACCTTGCCAT	128
Db	153	GGATCCCAACCGATCCGCCCGTCA CGACCGCGCGGCTCCGCGCTCGATCCACGGT	212
Qy	129	CTTCGGCGTCTTGGCAGTTCCCGCAGCGCTTCCACTTCGACAGGTTCCGCCCGCAGGTA	188
Db	213	CTTCGCGCGGTTCGCGCGCACCGCGACGCGCTTCGCTTCGTCAGGTTTGTGTGAGGTA	272
Qy	189	CGGGAAGAGCTACGGATCGGAGGAGGATCAAGAAGAGTTTCGGGATCTTCGTGGAGAA	248
Db	273	CGGCAAGAGCTACGAGAGCGCGCGGAGGTCACAAAGCGGTTTAGGATCTTCTCCGAGAG	332
Qy	249	TCTAGGTTTATCCGGTCACTAATCGGAAGGATCTGTGTATACCTAGGAATCAACCA	308
Db	333	TTCAGCTGTCGCGCTCCACCAATCGGAAGGCGCTCTCTACCGGCTTCGGCATCAACCG	392
Qy	309	ATTCCGCACTGACCTGGGAGGAATTCGGGACCAATCGCTTGTGTGGCGGCAGAACTG	368

[illegible]

```
QY 543 GAGCAGCATCAATATCTGAACAGCAACTTGTGAGCTGTGCTCAGCATTAATCAATTTGG 602
D 612 GCCCATCTCTCTCTGAGCAACAGCTTATTAAGTGTGCTTGTGATTAATCAATTTGG 671
QY 603 ATGCAATGAGGCTTTGCCCTTCCAAAGCCCTTTGAATACGTTAAGTACAAATGAGGATCGA 662
D 672 ATGCAACGAGGAGCCCTCCATCCAGCCCTTTGAATACATCAATACATATGAGGCTTGA 731
QY 663 CACAGAAACAGATTAATCAATCTTGGTGTGATGATGATGATGATGATGATGATGATGAT 722
D 732 CACTGAGGAATCTTACCCCTTACCAAGGTGTCAATGGAATCTGCAAGTTTAAAGAAATGAA 791
QY 723 TGTGCTGTCAAGGTCATTTGATTCGATAAATCAATCAACCTGGGTGCTGAGGATGATGAA 782
D 792 TGTGAGTCAAGTTTGGACTCGTTTAAATCAATCAACCTGGGTGCTGAGGATGATGAA 851
QY 783 GATGCAAGTGGCTTTGGTGGCTTACAGTGTGATGATGATGATGATGATGATGATGATGAT 842
D 852 GATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 911
QY 843 TCTGTACAGAAAGTGTATACAGCAGTGCACCTGTGGAAGAGATCAATGATGATGAA 902
D 912 GCTGTACAGAGCGAGTTTACATAGGCAATGTTGGAATCAACGATGATGATGAA 971
QY 903 CCACGAGTCTTGTGCGTGGTATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 962
D 972 CCACGCTGTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1031
QY 963 GAACATGAGGTTACAAATTTGGGTGACATGGCTACTTTAAGATGGAATCGGCAAGAA 1022
D 1032 GAACATGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1091
QY 1023 CATGTGTGCTGTGCAATTTGGCATCTTATCCCATTTGCTG 1064
D 1092 CATGTGCGGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1133
```

RESULT 6

```
US-10-425-114-5145
; Sequence 5145, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53)13B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 5145
; LENGTH: 1367
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700444833_FLI
US-10-425-114-5145
```

Query Match 51.7%; Score 553.8; DB 12; Length 1367;

Best Local Similarity 72.9%; Pred. No. 1.6e-175; Matches 743; Conservative 0; Mismatches 267; Indels 9; Gaps 2;

```
QY 55 GCGCTCGCGGTTTCGAAGAAGCAATCCAAATCCGCTCGGTTCACAAAGCCCTGACT-- 112
D 1 GCCAACTCGGCTTCGCGGACTCCAAACCGATCCGCTGTGACCGACCGCGGCTCC 60
QY 113 -CGATTAGCTGCCATCTCGGCTGCTTGGCAGTGTGCGCCACGCTTCCATTCGCA 171
D 61 GCGCTGAGTCCACGGTCTTCGCGGCTCGCGCGCACCCGCGAGCTGCTGCTTCGCA 120
```

RESULT 7

```
US-10-219-220-107
; Sequence 107, Application US/10219220
; Publication No. US2003008274A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: Death and their use in the modification of plant development
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
```

```
QY 172 CGGTTCCGCCGAGGTACGGGAAGAGCTACGATCGAGGAGGAGATCAAGAAAGAGTTTC 231
D 121 CGCTTCCGCCGAGGTACGGGAAGAGCTACGAGAGCGCGGAGGTCCATAAAGCGTTTC 180
QY 232 GGGATCTTCTGTGAGATCTAGCGTTTATCCCGTCCACTTAATCGGAGGATCTGTGAT 291
D 181 AGGATCTTCTTCCGAGAGCTCCAGCTGGTCCGCTCCCAACCGCAAGAGGCTCTCTCTAC 240
QY 292 ACCCTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGGAATTCGGGACCAATTCGCTT 351
D 241 CGCTCGGATCAACCGCTTCGCGGACATGAGCTGGAGGAGTTCGCTGCGACCGGCTC 300
QY 352 GGTGCGCGGAGAGACTGCTCGGCGAGCTGCGCATGGAACCCACCGTT-----TGTGAT 405
D 301 GCGCGAGCCGAGAACTGCTCGGCGAGCTTACCGGCAACCGGATGCGCGCGCGCC 360
QY 406 GGGTGTCTTCTCTAAACGAGGATTCGAGGAGCAAGGATAGTAGCCCTGTAAAGGAC 465
D 361 GTTCCGCTGCGGAGAGCAAGACTCGGAGGAGTGGATTTGTGAGCCAGTGAANAAC 420
QY 466 CAAGGAGCTGTGATCTTGTGGAATTTCACTTCACTTCTGAGCACTAGAGGCTGCATAT 525
D 421 CAGGCGCACTGTGATCATGCTGGAACCTTCAGCACTACTGCTGCACTTGGAGGCTGCATAT 480
QY 526 ACACAGCTAATCTGGAAGAGCAGCATCATTTATCTGAACAGCACTTGTGAGCTGTGCTCA 585
D 481 ACCAGGCACTGCGAAGCCATCTCTCTCTGAGCAACAGCTTGTGAGCTGTGCTT 540
QY 586 GCATTAATCAATCTTGCATGCAATGAGGTTTGCCTTCCCAAGCTTTGAATACGTTAAG 645
D 541 GCATTAACAAATTTCCGATGCAACGAGGCTTCCATCCGAGGCTTTGAATACATCAAA 600
QY 646 TACAATGGAGGATCGACACAGAGACTTATCCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 705
D 601 TACAATGGTGGCTTGCACACTGAGGAATCTTACCCCTTACCAAGGCTGCTGAATGGAATCTGC 660
QY 706 AACTTCAAGCAGGAGAAATGTTGTTCAAGGTCATTTGATTCGATAAATCAATCACCTGGGT 765
D 661 AAGTTTAAAGAAATGAGAAATGTTGAGTCAAGTTTGGACTCGGTTAAATCAATCACCTGGGT 720
QY 766 GCTGAGGATGAGTTGAAGCATGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
D 721 GCTGAGGATGAATGAAGGATGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 826 GTTGTCAAAAGTTTCAATCTGTACAAAGAAAGGTTATACAGCAGTGCACACCTGTGCAAGA 885
D 781 GTGATCACTGTTTCAAGCTGTACAGAGCGGAGTTTACACTAGCGACCATTTGTGGAAT 840
QY 886 GATCCAAATGATGTGAACCAACGAGTTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
D 841 ACACCGATGATGTGAACCAACGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 946 CCTTATTGGCTCATCAAGAACTCATGGGCTCAAAATTTGGGTCACAAATGGCTACTTTAAG 1005
D 901 CCTTCTGCTCATCAAGAACTCATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 1006 ATGGAATCGGCAAGAAACATGTGTGCTGTTGCAACTTGGCACTTCTTATCCCATTTGGC 1064
D 961 ATGGAATCGGCAAGAAACATGTGTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1019
```



```
Db 475 GTTAAAGATCAAGGAGCTGGGATCATGTCGACATTCAGACAACTGGGCTTTAGAA 534
Qy 517 GCTGATATACACAGCTAACTGGAAAGAGACATCATTTATCTGAACAGCACTTGTGGAC 576
Db 535 GCAGCCTATGCACAGCACTTTGGGAAGAGTATCTCTCTTCTGAGCAGCACTAGTGGAT 594
Qy 577 TGTGCTCAGCACTTCAATAACTTTGGATGCAATGAGGTTTGGCTTCCCAAGCCTTTGAA 636
Db 595 TGTGCTGCCCGTTTCAACACTTTGGCTGTAATGGTGGTTGCCATCAACAGCCTTTGAG 654
Qy 637 TACGTTAGTACAACTGGAGGCTACACACAGAAAGAGCTTATCTTATCTTGGTGTGCAAT 696
Db 655 TACATCAATACAACTGGAGGCTTGGAGCAGAGGAGCATACCCCTACACTGGAAAGAT 714
Qy 697 GGTATCTGCAACTTCAAGCAGGAGATGTTGGTGTCAAGGTCATTTGATTCGATAAACATC 756
Db 715 GGTGTATGCAAACTTCTAGCTGAAATGTTGCCGTTCAAGTCAATGACTCGGTCAATATC 774
Qy 757 ACCCTGGTGTGAGGATGAGTTGAAGCATGAGTGGGCTTGGTGGTCCAGTTAGCGTT 816
Db 775 ACCTTGGGTGTGAGGATGAGTTGAAGCATGAGTGGGCTTGGTGGTCCAGTTAGCGTT 834
Qy 817 GCATTTGAGTGTGAGGATGAGTTGAAGCATGAGTGGGCTTGGTGGTCCAGTTAGCGTT 876
Db 835 GCCTTTGAGTGTGAGGATGAGTTGAAGCATGAGTGGGCTTGGTGGTCCAGTTAGCGTT 894
Qy 877 TGTGAAGAGATCCAACTGGAGTGAAGCATGAGTGGGCTTGGTGGTCCAGTTAGCGTT 936
Db 895 TCGGCTAGCACTTCCAGGATGTGAACCATGCGCTTCTGTTGGGCTATGGAGTTGAA 954
Qy 937 GACGGATTCCTTATGCTGCTCATCAAGAACTCATGGGTGACAAATGGGTTGCAATGGC 996
Db 955 AATGGCGTCCCATATTTGCTCATATAAAATTCATGGGAGAAAGTTGGGTTGAGATGGC 1014
Qy 997 TACTTTAAGATGGAATCGGCAAGAACATGTTGGTGTGCAACTTGGGCTATCTTATCCC 1056
Db 1015 TACTTCAAAATGGAATCGGAGAACATGTTGGTGTGCAACTTGGGCTATCTTATCCA 1074
Qy 1057 ATTGTGC 1064
Db 1075 GTTGTGC 1082
```

RESULT 12

```
US-10-424-599-122629
; Sequence 122629, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 122629
; LENGTH: 1933
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1933)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_81740C.1
US-10-424-599-122629
```

Query Match 45.0%; Score 481.8; DB 12; Length 1933;
Best Local Similarity 69.2%; Pred. No. 4e-151;

```
Matches 671; Conservative 0; Mismatches 297; Indels 1; Gaps 1;
Qy 97 ACACAAAGGCTGACTCGATTGAGCTGCCATCCTCGGGTCTTTCGACAGTTGCCGCCAC 156
Db 234 ATACGAATGTGCGGGGTGGAGCGGAGTGGTTCGGGTGATCGGGCAGTGCCTGGCGC 293
Qy 157 GCCTTCCACTTCGACCGTTTCGCCCGCAGGTACGGGAAGAGCTTACCGATCGGAGGAGG 216
Db 294 GCCTGAAAGTTTCTAGGTTTCATGACGAGTTCGGGAAGAGTTTACCGAAGCGAGGAG 353
Qy 217 ATCAAGAGAGGTTTCGGGATCTTCGTTGGAGATCTAGCGTTTATCCGTCCTCAATCG 276
Db 354 ATGAGGAGAGGTTACGAGATATTTCTCGAAACCTCAGGTTTATCCGTTCCCAACAAG 413
Qy 277 AAGGATCTGTGCTATACCTTAGGAATCAACCAATTCGCCGACCTGAGTGGGAGGAATTC 336
Db 414 AACCGCTTGCCTTACACTCTCTCTGTTAATCAATTTTCTGATTGGACTTTGGGAGGATTC 473
Qy 337 CGACCAATCGCTTTCGGTGGCGGAGAACTGCTCGGCACTGCGGATCGGAACCAACCG 396
Db 474 AAAAGACACAGACTAGCGCTGCCAAATTTGCTCTGCCACTTTTAAACGCAACCAAG 533
Qy 397 TTTGTGATGGCGTGTCTTCTGTAACGAGGATTTGGAGGAGCAAGGAGTAGTGAGCCCT 456
Db 534 CTCACTGATGCTGTTCTTCTCCCAAGAACTGGAAGAAAGAAATAGTGTGAT 593
Qy 457 GTAAAGAACCAAGAAAGTGTGATCTTGTGTGACTTTTCAGTACTACTGAGCAGCTAGAG 516
Db 594 GTTAAAGATCAAGGAGCTGCGGATCATGCTGGACATTCAGCAACAATCGGGCTTTAGAA 653
Qy 517 GCTGCATATACACAGCTTAACTGGAAGAGACATCATTTCTGAAACAGCAACTTTGTGGAC 576
Db 654 GCAAGCTATGCAACAGCTTTGGAAAGATATCTCTTCTTCTGAGCAGCAGCTAGTGGAT 713
Qy 577 TGTGCTCAGCACTTCAATAACTTTGGATCAATGGAGTTTGGCTTCCCAAGCCTTTGAA 636
Db 714 TGTGCTGGCCTTTCAACAACTTTGGCTGTAAATGGTGGTGGTGCATCAACAAGCCTTTGAG 773
Qy 637 TAGC-TTAAGTACAATGGAGGATTCACACAGACAGACAGCTTATCCATACCTTGGTCAA 695
Db 774 TACATTTAAATACAATGGTGGACTAGACAGAGGAGCAATATCCCTACACAGGAAAGA 833
Qy 696 TGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCATTTGATTTCGATAAACA 755
Db 834 TGTGCTGCAAAATTCAGCTGAAATGTTGCTGTTCAAGTCTTCTGACTCTGTAATAT 893
Qy 756 CACCTGGGTGCTGAGGATGAGTTGAAGCATGACAGTGGGCTTGGTCCGTCAGTTAGCGT 815
Db 894 CACCTTGGGTGCTGAGAAATGAATTTAAACATGACAGTTGCATTTGTTTCGGCCGTTAGTGT 953
Qy 816 TGCATTTGAGGTTGTGAAAGGTTTCAATCTGTACAAAGAAAGGTGTATACAGCAGTGACAC 875
Db 954 GGCCTTTCAGTGTGAAATGGGTTCCATTTCTACGAGAAATGGAGTTTACACTAGTGCAT 1013
Qy 876 CTGTGAAGAGATCCAAATGATGTGAACCAAGCAGTTCTTGGCTCGGTTATGAGTTCGA 935
Db 1014 TTGCGGTAGCACTTCCAGGATGTGAACCATGCGCTCTTGTGTGGGGTATGGAGTTGA 1073
Qy 936 GGAAGGATTCCTTATTTGGCTCATCAAGAACTCATGCGGTACAAATTTGGGTGCAATGG 995
Db 1074 GAATGGGCTCCCATATTTGGCTCATATAAAATTCATGGGAGAAAGTTGGGTGAGATGG 1133
Qy 996 CTACTTTAAGATGAACTCGGCAAGAACATGTTGTTGTGCAACTTGGGCTCTTATCC 1055
Db 1134 CTACTTTCAAAATGAAATGGGAGAACATGTTGTTGTGTTGCAACTTGTGCGTCTTATCC 1193
Qy 1056 CATTGTGC 1064
Db 1194 AGTTGTGC 1202
```

RESULT 13

US-09-770-445-20/c

Qy	543	GACGACATCATTTCTGAACAGACAACTTGGGACTGTGCCTCAGCATTCAAATCACTTTGG	603
Db	734	AGAAATATCTTTGTCGAGGCAACAGCTTGTGGAAATGTGCTGGTACTTTCATTAACITTTGG	675
Qy	603	ATGCAATGGAGGTTTGCTTCCCAAGCCTTTGAATACGTTAAGTACAAATGGAGGCATCGA	662
Db	674	TTGTGATGGTGGACTTCCCTTCTCAAGCCTTTGAATACATTAATAACACGCTGGGCTCGA	615
Qy	663	CACAGAACACATTAATCCATACCTTGGTGTCATGGTATCTGCAACTTCAAGCAGGAGAA	722
Db	614	CACGGAGAGGCTTATCCTTACACCGGAAAGACGGTGGCTGCAAAATTTTCAGCGCAAAA	555
Qy	723	TGTTGGTGTCAAGGTCAATTAATTCGATTAACATCACCTCGGTGCTCAGGATGATTTGAA	782
Db	554	CATCGGTGTACAAGTCGGTGACTCTGTCAACATTAACCTGGGTGCAGAAGATGAACGAA	495
Qy	783	GCATGCAGTGGGCTTGGTGGCTCCAGTTCAGTTGCAATTTGAGTTCTGAAAGGTTTCAA	842
Db	494	GCACGGCTCGGGTGGTAAAGCCAGTGAAGTGTGGCGTTTGTAGAGTTGTATCATGAATTCAG	435
Qy	843	TCGTGTCAAGAAAGGTGTATACAGCAGTGTACACCTGTGGAAGAGATCCAAATGGATGTGAA	902
Db	434	GTTTTATAAGAGGGAGTTTTTACTAGCAATACATGTGGTAAACACTCCAATGGATGTAAA	375
Qy	903	CCAGGCAGTTCTTGGCCGTCGGTTATGGAGTCGAGGACGGATTCCTTTATTTGGCTCATCAA	962
Db	374	CCATGCAGTGTATAGCAGTTGGTTATGGAGTTGAAGACGACGCTCCGCTACTGGCTTTATAA	315
Qy	963	GAACTCATGGGTACAAATTTGGGGTGACAATGGCTACTTTTAAAGATGAAACTTCGGCAAGAA	1022
Db	314	GAACTCATGGGAGGTGAATGGGGAGACAAATGGCTACTTCAAGATGAAATGGGAAGAA	255
Qy	1023	CATGTGTGGTGTGCAACTTGGCGCATCTTATCCCATTTGTGGC	1064
Db	254	CATGTGTGGTGTGCAACATGTTTCATCGTATCCGGTTGTAGC	213

RESULT 14
US-10-219-220-241
; Sequence 241, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OP INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of plant development
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 241
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Pinus radiata
US-10-219-220-241

	Query Match	38.0%;	Score 407.4;	DB 14;	Length 1652;
	Best Local Similarity	65.7%;	Pred. No. 4.6e-126;		
	Matches 626;	Conservative 0;	Mismatches 321;	Indels 6;	Gaps 2;
Qy	116	TTGAGCTGCGCATCCTCGGGCTCTTGCGCAGTTGCGGCCACGCCCTTCACCTTCGCACGGT	175		
Db	385	TGGAGTCCAATTGGTTAAATCTTGGGAACCAATCCAAAGTCTCTACAATTTGCGGAGT	444		
Qy	176	TCGCCCCAGGTACGGGAAGAGCTACCGATCGGAGGAGGATCAAGAGAGGTTTCGGGA	235		
Db	445	TCGCTGTCAAGATATGGCAAGAGGTACCATCTGTCCATTCAGCTTGTGCATAGATTCATGT	504		
Qy	236	TCCTTCGTGGAGAATCTACGCTTTTATCGGTTCACTAATCGGAAGGATCTGTCGTATACCC	295		

Db 505 TCTTGTGAAGAACTGGAGCTGATCGAGTCAAGAAACAGAAATGAAGCTTCCTTATACTT 564
Qy 296 TAGGAATCAACCAATTCGCGACTGACCTGGAGGAATTCGGACCAATCGCCTTGCTG 355
Db 565 TGGCAATAAATGAGTTGCTGACATAACATGGAGGAATTCATGGACAAATATTGGGTG 624
Qy 356 CGGCGCAGAACTGCTGGCGACTGGCATGGAACACCGGTTGTCGATGGCTGCTTC 415
Db 625 CTTACAGAACTGTTGCG---CTACCCACAGTAACCAATGAGTGAAGTATGCCAGCTTC 681
Qy 416 CTGTAAACAGGGAATGGAGGAGCAAGGATAGTGGCCCTGTAAAGACCAAGGAAGCT 475
Db 682 CTGCGAAGAAAGACTGGAGACAAGAAAGCATAGTGAGTCTGTAAAGAAACCAAGCCCAT 741
Qy 476 GTGGATCTGTGGACTTTCAGTACTAGTGGACACTAGAGCTGCATATACAGACTAA 535
Db 742 GTGGATCTGTGGACATTCAGCACTACTGAGCACTAGAGGCTGCCTATACTCAGGCTA 801
Qy 536 CTGGAAGAGCAGACATATATCTGAACAGCAACTTGTGCACTGTGCTTCAGCATCAATA 595
Db 802 CAGGAAGACTGTTATCTGTCTGAACAGCAGCTGTTGACTGTGCTGGAGCAATTAACA 861
Qy 596 ACTTTGGATGCAATGGAGGTTTCCTCCCAAGCCTTTGAATAGCTTAAAGTACAAATGGAG 655
Db 862 ACTTTGGTGTCAATGGTGAAGTGCCTCCCAAGCAATTTGAGTATATCAAGTACAAACGGAG 921
Qy 656 GCATCGACACAGACAGACTTATCCATACCTTGTGTCATGTTGTCATGCTGCAACTTCAAGC 715
Db 922 GCTTGACACTGAGGAAGCTATCTTACAGCTAAAGATGGTGTGTTGTAATATGATG 981
Qy 716 AGGAATATGTTGTTCAAGTCAATGATTCGATAAACAATCACCCCTGGTGTGCTGAGGATG 775
Db 982 TAAACAATGTCGTTGTAAGTGTGCTGATAGTGTCAACATCAGTTTGGGGGCAGAAGATG 1041
Qy 776 AGTTCAAGCAGTGGGCTTGGTGGTCCAGTTAGGCTTGGATTTGAGTTGCTGGAAG 835
Db 1042 ACTCAAGACTGCTGGGCTTAGTTCGCCCTGTGAGTGGCATTCAGTTAATTCGG 1101
Qy 836 GTTTCATCTGTACAAGAAAGTGTATACAGCAGTGACACCTGTGGAAGAGATCCAATGG 895
Db 1102 AATTTTCGATTTTACAAGGAAGAGTCTTTTACAAGCACAAGTTGTGTTCAAGGCCCAATGG 1161
Qy 896 ATGTGAACACGCACTTCTGCGCTGCTGATGAGT---CGAGGAGGGATTCCTTAT 952
Db 1162 ATGTCAACCATGCTCTTTTGGCCGTGGGTATGTTGTTAGTGGAGGGGACTCCACAT 1221
Qy 953 GCTCATCAAGAACTCATGGGGTACAAATTTGGGGTGACAATGGGCTACTTTAAGATGGAAC 1012
Db 1222 GATCATCAAGAAATTCCTGGGAAGAGAGCTGGGTGTTGATGGATCTTCAAGATGGAGT 1281
Qy 1013 TCGGCAAGAACATGTGGTGTGTGCAACTTGGCAATCTTATCCCATTTGGGCT 1065
Db 1282 TAGGGAAGAATATGTGTGTGTGGACTTGTGCTTGTGCTTGTGCTTATTTGTTCT 1334

RESULT 15

US-10-425-114-3201
; Sequence 3201, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3201

; LENGTH: 830
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700243475_FLI
US-10-425-114-3201

Query Match 34.5%; Score 369; DB 12; Length 830;
Best Local Similarity 77.5%; Pred. No. 2.8e-113; Indels 0; Gaps 0;
Matches 447; Conservative 0; Mismatches 130;
Qy 488 GGAATTTTCACTACTACTGGAGCACTAGAGCTGCATATACAGCTTAACCTGGAAGAGCA 547
Db 2 GGACCTTCACACTACTTGTGTGCACTTGGAGCTGCATATATACCCAGGCACTGGCAAGCCCA 61
Qy 548 CATCATTTCTGAACGAACCTTGTGGAGCTGCTCAGCATTCATTAACCTTTGGATGCA 607
Db 62 TCTCTCTCTCTGAGCAACAGCTTATTGACTGTGGTTTTCATTCAACAAATTTCCGATGCA 121
Qy 608 ATGGAGGTTTGCCTTCCCAAGCCTTTGAATACCTTAAAGTACATGAGGAGCATCGACACAG 667
Db 122 ACAGAGGCTTCCATCCAGGCCCTTTGAATACATCAATCAATGCTGCTTGCACATG 181
Qy 668 AACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTG 727
Db 182 AGGAATCTTACCTTACCAAGGTGTCAATGGAATCTGCAAGTTTAAAGAAATGAGAAATGTTG 241
Qy 728 GTCTCAAGGTTCATTGATTCGATAAACATCACCTGGGTGCTGAGGATGAGTTGAAGCATG 787
Db 242 GATCAAGGTTTGGACTGGTTAATCACCTGGGTGCTGAGGATGAGTTGAAGCATG 301
Qy 788 CAGTGGGCTTGGTGGCTTCCAGTTAGCGTTGCAATTTGAGGTTGTGAAGGTTTCAATCTGT 847
Db 302 CTGTGCTGTGGTTCGCCAGTTAGTGTGCTTCCAGGTGATCACTGGTTTCAGGCTGT 361
Qy 848 ACAAGAAAGGTGTATACAGCAGTGACACCTGTGGAAGAGATCCAATGGATGTGAACACG 907
Db 362 ACAAGAGCGAGTTTACACTAGGCACTTGTGGAACCTACCGATGGATGTGAACACG 421
Qy 908 CAGTCTTTCGCGTCCGTTATGAGTGGAGCAGGGAATTCCTTATTTGGCTCATCAAGAACT 967
Db 422 CTGTTCTGGCTGTTGGCTACGGTGTGGAAGATGGTGTACCTACTGGCTCATCAAGAACT 481
Qy 968 CATGGGGTACAAATTTGGGTGCAATGGCTACTTTAAGATGGAACCTCGGCAAGAACTGT 1027
Db 482 CATGGGGCGCTGACTGGGGTGTAGGGGTACTTTCAAGATGGAATGGGCAAGAACTGT 541
Qy 1028 GTGGTGTGCAACTTGGCATCTTATCCCATTTGTGGC 1064
Db 542 GCGGTGTTGCTACGTGTGCATCTTATTTGTGCGC 578

Search completed: March 29, 2004, 22:09:43

Job time : 332 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 29, 2004, 19:15:42 ; Search time 2391 Seconds
(without alignments)
13376.159 Million cell updates/sec

Title: US-10-087-714-1
Perfect score: 1071
Sequence: 1 atggcagctaaagctctctt.....atccattgtggtgtgtag 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_estc.*
9: gb_est1.*
10: gb_est2.*
11: gb_est3.*
12: gb_est4.*
13: gb_est5.*
14: gb_est6.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_pbg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	571.2	53.3	1532	11	AV106223 Zea mays
2	434.6	40.6	989	14	CK263045 EST709123
3	430.8	40.2	829	14	CB341855 CA32EN000
4	417.8	39.0	1005	14	CF212812 CGF100065

5	410.8	38.4	814	12	BG594735
6	410	38.3	985	14	CK287181
7	408	38.1	731	13	BQ281381
8	407.8	38.1	748	14	CB982399
9	405.6	37.9	936	14	CK268548
10	403.4	37.7	920	14	CK257583
11	403.8	37.6	848	14	CF510505
12	401.8	37.5	886	14	CF510584
13	398.6	37.2	977	14	CK295860
14	394.2	36.8	720	10	BE131652
15	390.8	36.5	965	14	CK294622
16	389.8	36.4	910	13	BQ797116
17	387.8	36.2	794	10	BE195255
18	386.2	36.1	768	14	CB892658
19	384.8	35.9	701	14	CB894684
20	383.8	35.8	653	14	CB917819
21	383.4	35.8	924	14	CK294694
22	378.6	35.4	755	14	CB981790
23	378.6	35.4	930	14	CK295057
24	378.4	35.3	1230	10	BE422216
25	378	35.3	709	12	BG155091
26	377	35.2	852	14	CK263154
27	375.2	35.0	639	10	AW155820
28	375.2	35.0	665	10	AW399564
29	375	35.0	863	29	CG789556
30	373.6	34.9	760	10	BF267048
31	372.6	34.8	894	13	BU039998
32	370.8	34.6	609	14	CB006195
33	370.6	34.6	827	9	AA979924
34	369.6	34.5	901	14	CK298522
35	368.4	34.4	635	13	BU039803
36	367.8	34.3	946	14	CK285936
37	367.4	34.3	710	12	BI267420
38	366.6	34.2	703	12	BI269594
39	366.2	34.2	815	14	CA809544
40	365.4	34.1	648	13	BQ625193
41	365.2	34.1	888	14	CF519040
42	364.8	34.1	911	14	CK286268
43	364.4	34.0	888	14	CK292918
44	363.8	34.0	755	12	BG645271
45	362.6	33.9	833	14	CB666317

ALIGNMENTS

RESULT 1	AY106223	1532 bp	mRNA	linear	HTC 16-OCT-2002
LOCUS	AY106223	Zea mays	PCO114759	mrna	sequence.
DEFINITION	AY106223	HTC			
ACCESSION	AY106223.1	GI:21209301			
VERSION	AY106223.1	GI:21209301			
KEYWORDS	HTC				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	1 (bases 1 to 1532)				
AUTHORS	Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.				
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2 (bases 1 to 1532)				
AUTHORS	Coe, E.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the				

2d. Set 4 were grown under the standard conditions and then were heat stressed by placement at 35 C. Heat stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, 2d and 4d and heat-stressed roots were harvested at 6 hr, 12 hr, 1 d, and 4d. RNA was isolated from all tissues and equal RNA from each tissue and stress was pooled to construct the cDNA library. RNA sample."

ORIGIN

Query Match 40.6%; Score 434.6; DB 14; Length 989;
 Best Local Similarity 68.5%; Pred. No. 2.9e-94;
 Matches 599; Conservative 0; Mismatches 276; Indels 0; Gaps 0;

QY 107 CTGACTCGATTGAGCTGCTCCATCTCTCGCGTCTTGGCAGTTCGCCACGCCCTTCCACT 166
 DB 115 CCGAGAGCTGAGAACGGAATCTTCAAGTCTCGGCAGACTCGCAATGCTCTCTCT 174
 QY 167 TCGACGGTTCCCGCGAGTACGGAGAGAGCTACGGATCGGAGGAGAGATCAAGAGA 226
 DB 175 TCGCTCGCTTTGCTATCAGGATCGGAAAGGTACAGTCCGTTGAGGAGATCAAGCAA 234
 QY 227 GGTTCGGGATCTTCGTGGAGATCTAGCGTTTATCCGCTCCACTAATCGAAGGATCTGT 286
 DB 235 GGTTCGAGATATTTTGGCAATCTGAAGATGATCCGATCGCATACAGCAAGGACTAT 294
 QY 287 CGTATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGAGGAATTCGGGACCAATC 346
 DB 295 CATACAAACTCGGTGTCAATGAGTTTACGACCTAACATGGGATGAGTTCGGTAGACACA 354
 QY 347 GCCTTGGTGGCGGAGACTCTCGGCGACTGCGCATGCGCATGGAACACCGGTTGTCGATG 406
 DB 355 AGTTGGGGCATCTCAAAACTGTTCTGCCACTACAAAGGGCAATCTCAAGCTAATACG 414
 QY 407 GGGTGTCTTCTTAACGAGGGATTTGGAGGAGCAAGGGATAGTGAGCCCTGTAAAGGACC 466
 DB 415 TTGTTCTGCGACAGAGGAGACTGGAGGAGATGGTATTGTTAGCCAGTGAAGGCAC 474
 QY 467 AGGAAGCTGTGATCTTCTGACTTTCAGTACTACTGGACACTAGAGCTGCATATA 526
 DB 475 AGGGCAAGTGGGATCTTCTGCGACATTCAGCACTACTGTTGTCATAGGAGCAGCATATG 534
 QY 527 CACAGTAACTGGAAGACACATCATATCTGAACAGCAACTTGTGACTGTGCTCAG 586
 DB 535 CCGAGCATTTGGAGGAACTCTCTGTCAAGAGCAGCAGCTTGTGACTGTGCTGGAG 594
 QY 587 CATTAATACTTTGATGCAATGAGGTTTGGCTTCCCAAGCCTTTGAATACGTTAAGT 646
 DB 595 CTTTAAATACTTTGGCTGCAATGGGGGTTGCCATCACAGCCTTTGAGTACATTAAT 654
 QY 647 ACAATGGAGCATCGACACAGACAGACTTATCCATACCTTGTCTCAATGTATCTGCA 706
 DB 655 TCAATGGTGTCTTGACACTGAAGAGCATATCCATACACCGGCAAGATGGCATATGTA 714
 QY 707 ACTTCAAGCAGGAGATGTTGGTGTCAAGGTCAATTCGATTAACATCAACCTGGGTG 766
 DB 715 AATTTCTACAAGCAATATTTGGTGTCAAAGTCATCAGTTCTGTCAATATTACCTGGGTG 774
 QY 767 CTGAGATGATGTTGAAGCATGAGTGGGCTTGGTGGCTCCAGTTAGCGTTGCAATTTAGG 826
 DB 775 CTGAAGATGAATGAATAACGAGTTGCAATGGTGGCTGTTAGTGTGCTTTTGGAG 834
 QY 827 TTGTGAAGGTTTCAATCTGTACAGAAAGGTGTATACAGCAGTGCACACTCTGGAAGAG 886
 DB 835 TGGTAAGAAGTTTCAACAGATATAAGAGCGGAGTTTACACAGCACTGAATGTGGGACA 894
 QY 887 ATCCAAATGAGTGAACCAACGAGTTCTTGGCGGTTCGGTTATGGATCGAGGAGGATTC 946
 DB 895 CTCCCATGGACGTAACCACTGCTGTTCTTGGTGTGGGGTACGGTGTGTAATATGGCGTC 954
 QY 947 CTTATTGGCTCATCAAGAACTCATATGGGTACCAAT 981
 DB 955 CCTACTGGCTCATTAAGAACTCATATGGGGAGCAGAT 989

RESULT 3

CB341855 829 bp mRNA linear EST 10-APR-2003
 LOCUS CA32EN0002_IIIBF_H04 Cabernet Sauvignon Leaf - CA32EN Vitis
 DEFINITION vinifera cDNA clone CA32EN0002_IIIBF_H04 5', mRNA sequence.

ACCESSION

CB341855

VERSION

CB341855.2 GI:29785118

KEYWORDS

EST.

SOURCE

Vitis vinifera

ORGANISM

Vitis vinifera
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; Vitaceae; Vitis.

REFERENCE

1 (bases 1 to 829)
 Goes da Silva F., Lim H., Iandolino A., Baek J., Jones K.,
 Walker M.A. and Cook D.R.

AUTHORS

Transcriptional responses of Vitis vinifera to infection by the
 bacterial pathogen Xylella fastidiosa

TITLE

Unpublished (2003)

JOURNAL

On Mar 14, 2003 this sequence version replaced gi:28962822.

COMMENT

Contact: Douglas Cook, PhD
 CAES Genome Facility
 UC Davis, Plant Pathology
 One Shields Ave, Davis, CA 95616, USA
 Tel: 530 754 6561
 Fax: 530 754 6617
 Email: drcook@ucdavis.edu

Seq primer: ACGTACCGGACATATGCC.

FEATURES

source

Location/Qualifiers

1..829

/organism="Vitis vinifera"

/mol_type="mRNA"

/cultivar="Cabernet Sauvignon"

/db_xref="taxon:29760"

/clone="CA32EN0002_IIIBF_H04"

/sex="Hermaphrodite"

/dev_stage="Mid-season leaf material, collected July 25, 2001."

/lab_host="DH5alpha"

/clone_lib="Cabernet Sauvignon Leaf - CA32EN"

/notes="Organ: Leaf; Vector: pDNR; Site: 1: SfiI; Site 2:
 SfiI; CA32EN is a cDNA library of Cabernet Sauvignon
 leaves. The leaves were collected on July 25, 2001, in
 Napa Valley, California, and represent leaves in
 mid-season development. These leaves were asymptomatic and
 verified to be non-infected with the bacterial pathogen,
 Xylella fastidiosa, based on a diagnostic assay using PCR
 and Xylella-specific primer pairs. cDNA made by oligo-dT
 priming and directionally cloned. 5' and 3' adaptors were
 used in cloning as follows:
 5'-AAGCAGTGGTATCAAGCAGATGGCCATTACGCCGGG-3' and
 5'-ATTAGAGCGCGGCGGCGGACATG-dt(30)NN-3'. Library was
 constructed using the Clontech Creator SMART kit and
 size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Query Match 40.2%; Score 430.8; DB 14; Length 829;
 Best Local Similarity 70.7%; Pred. No. 2.2e-93;
 Matches 573; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 125 CCATCTCTCGCGTCTTGGCAGTTCGCCACGCTTCCACTTCGCAAGTTCGCCCGCA 184
 DB 3 CCGTCTCTCGCGTCTTGGCAGTTCGCCACGCTCCTCTTCCGAGCTTCGCTCACA 62
 QY 185 GGTACGGGAGAGCTACGGATCGGAGGAGAGATCAAGAGAGTTCGGGATCTTCGTGG 244
 DB 63 GGTATGGGAGAGATTCAAGCGGTGACGAGATTAACTGAGATTCAGATTTTCGCG 122
 QY 245 AGAATCTAGGCTTTATCCGCTCCACTAATCGAAGGATCTGTGATATACCTAGGAATCA 304
 DB 123 AGAATTTGAAACTCATCAGATCCACCAACAGAAAGGCTTGCCTTATCTACTAGCTGTA 182

QY 725 TTGGTCTCAAGTTCATTGATTCGATAAATCAATCACCTCGGTGCTGAGGATGAGTTGAAGC 784
 Db 808 TTGGTCTCAAGTTCCTCGACTCTGTGAATATTACCTCGGTGCTGAGGATGAGTTAAAGC 867
 QY 785 ATGCACTGGGCTTTGGTGGCTGCGTTCAGTTCAGTTAGCGTTGCAATTCAGGTTGTGAAAGGTTTCAATC 844
 Db 868 ATGCACTGGCAATTTGTTCTGTCCTCAGTGTGAGTGTGCAATTCAGGTTGTCCATGATTTCCGAT 927
 QY 845 TGTACAGAAGGTGTATACAGCAGT-GACACCTGTGGAGAGATCCCAATGATGTGAAC 903
 Db 928 TTACGATCAGAGGTTTACAGAAGTGGAACTTTGGCAGCAGTCCCAATGATGTGAAC 987
 QY 904 CACGCACTTCTTGCCTGC 921
 Db 988 CATGCTGTTCTTGCACTC 1005

RESULT 5
 BG594735
 LOCUS EST493413 cSTS Solanum tuberosum cDNA clone CSTS8E24 5' sequence,
 DEFINITION mRNA sequence.
 ACCESSION BG594735
 VERSION BG594735.1 GI:13612875
 KEYWORDS EST.
 SOURCE Solanum tuberosum (potato)
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum.
 van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,
 Bougri,O., Buell,C.R., Roming,C., Tanksley,S. and Baker,B.
 Generations of ESTs from sprouting potato eyes
 Unpublished (2000)
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@igr.org
 This clone can be obtained from the University of Arizona Genomics
 Institute. Orders can be made through URL:
 http://genome.arizona.edu/orders/
 Seq primer: M13F-R.

FEATURES
 Location/Qualifiers
 1..814
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="CSTS8E24"
 /tissue_type="sprouting eyes from tubers"
 /dev_stage="12-14 weeks post harvest"
 /lab_host="SOLR"
 /clone_lib="cSTS"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
 taken from tubers. The tubers were incubated at 26C in the
 dark for 2-3 weeks prior to sprouting. The eyes were
 frozen in liquid nitrogen immediately upon removal from
 tubers."

ORIGIN
 Query Match 38.4%; Score 410.8; DB 12; Length 814;
 Best Local Similarity 70.0%; Pred. No. 1.5e-88;
 Matches 567; Conservative 0; Mismatches 242; Indels 1; Gaps 1;
 QY 165 CTTGCGACGCTTGGCCGCGAGTACGGAGAGCTACGGATCGAGGAGAGATCAAGAA 224
 Db 6 CTTGCGCTGCTTGTGATCAGCATCGGAAGGTTACGATCCGTTGAGGAGATCAAGCA 65
 QY 225 GAGGTTGCGGATCTTCGTGTGAATCTAGCGTTTATCCGGTCCCACTAATCGGAAGGATCT 284
 Db 66 AAGGTTGCGAGATATTTTGGCAATCTGAAGATGATCCGATCGCATACAGCAAGGACT 125

QY 285 GTCGTATACCTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGAAATTCGGACCAA 344
 Db 126 ATCATACAAATCTCGTGTCTCAATGATGTTTACCGACTTAACATGGGATGAGTTCCGTAGACA 185
 QY 345 TCGCTTTGGTGGCGCGCAGAACTGCTCGGACACTGCGCATGGAACACACCGTTTTCGA 404
 Db 186 CAAGTTGGGGGCACTCTCAAACTGTTCTGCCATCAAAAGGGCAATCTCAAGCTAACTAA 245
 QY 405 TGGCGTCTTCTCTGTAAACGAGGGANTGGAGGGAGCAAGGATAGTGAGCCCTGTAAAGGA 464
 Db 246 CGTCTTCTGTCGACAGACGAGGACTGAGAGGAAGATGGTATTGTTTACCCAGTGAAGGC 305
 QY 465 CCAGGAAGCTGTGGATCTTCTGGACTTTTCAGTACTACTGGAGCACTAGAGGCTGCATA 524
 Db 306 ACAGGCAAGTGGGATCTTCTGGACATTCAGCACTACTGGTGCACTAGAGGAGCATTA 365
 QY 525 TACACAGCTAACTGGAAAGAGCACATCATTTATCTGAACAGCAACTTGTGGACTGTGCTC 584
 Db 366 TGCCCAAGCAATTTGGGAAGGGAATCTCTCTGTGAGAGCAGCAGCTTGTGGACTGTGCTGG 425
 QY 585 AGCAATTCAAATCTTTGGATGCAATGGAGGTTTGGCTTCCCAAGCCCTTGAATACGTAA 644
 Db 426 AGCTTTTAAATAACTTTTGGCTGCAATGGGGGTTGCCATCAAAAGCCCTTTCAGTACATTAA 485
 QY 645 GTACAATGGAGGCAATCGACACAGAAAGCACTTATCCATACCTCTGGTGTCAATGATCTG 704
 Db 486 ATTCAATGGTGTCTTGACACTGAAAGAGCATATCCATACACGGCAAGATGGCATATG 545
 QY 705 CAATTCACAGCAGAGAAATTTGGTGTCAAGGTCAATGATTCGATAAACATCAACCTGGG 764
 Db 546 TAAATTTCTCAAGACCAATATTTGGTGTCAAAGTCATCAGTTCTGTCAATATTACCCCTGGG 605
 QY 765 TGCTGAGGATGAGTTGAAGCATGCAGTGGCTTGGTCCGTCAGTTAGCGTTGCATTGGA 824
 Db 606 TCGTGAAGATGAATGAAATACGCAATTCGATTTGGTTAGGCCCTGTAGTGTGCTTTGA 665
 QY 825 GGTGTGAAGGTTTCAATCTGTCAAGAAAGGTGTATACAGCAGTGACACCTGTGGAAG 884
 Db 666 GGTGTTAAAGGTTTCAACAGTATTAAGAGCGGAGTTTACACAGCAGTCAATGTGGCGGA 725
 QY 885 AGATCCCAATGATGTGAACCAAGCAGTCTTTCGGCTCGGTTATGGAGTCGAGGACGGAT 944
 Db 726 CACTCCCATGGAGTAAACCATCTGTGCTGTGTGGCTCGGTGGTGTGAAAAATGCGCT 785

945 TCCTTATTGGCTCATCAAGAACTCATGGGG 974
 786 TCCTACTGCTCAT-AGAACTCATGGGG 814

RESULT 6
 CK287181
 LOCUS EST749903 Nicotiana benthamiana mixed tissue cDNA library,
 DEFINITION normalized, full-length Nicotiana benthamiana cDNA clone NEMB738 5'
 end, mRNA sequence.
 ACCESSION CK287181
 VERSION CK287181.1 GI:39863480
 KEYWORDS EST.
 SOURCE Nicotiana benthamiana
 ORGANISM Nicotiana benthamiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Nicotiana.
 Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A., Day,B.,
 Staskiewicz,B., Jin,H. and Baker,B.
 Generation of EST sequences from Nicotiana benthamiana
 Unpublished (2003)
 Other ESTs: EST749904
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA

TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give pBluescript phagemids before normalization was carried out. The mass excision of phagemid library and normalization were done in HT Nguyen lab by D. Zhang at Texas Tech University. Normalization protocol used was that of Soares'. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

```

Query Match      38.1%; Score 408; DB 13; Length 731;
Best Local Similarity 73.5%; Pred. No. 7e-88;
Matches 535; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 232 GGGATCTTTGCGGAGAACTACGCTTTATCCGGTCCCAATAACGGAAGATCTGTCTGAT 291
DB 4 GGGATCTTTCCGAGAGCTCGAGCTCGCTCGCTCCACCAACCGGAGGGCTCCCTAC 63

QY 292 ACCTAGGAATCAACCAATTCGCGACCTGACCTGGGAGGATTCGGACCAATCGCTT 351
DB 64 CGCTTCGGCATCAACCGTTTCGGGACATGAGCTGGGAGGATTCAGGCGAGCGGCTC 123

QY 352 GTGCGGCGCAGAACTGCTCGCGCACTCGCATGGAACCAACCGGTT---TGTCGATGCG 408
DB 124 GCGCGCGCGCAGAACTGCTCGCGCAGCTCGCGGCAACCAACCGGATCGCGCGCGCC 183

QY 409 GTGCTTCTGTAAACGAGGATTTGAGGAGGAGCAAGGATAGTAGAGCCCTGTAAAGAGCAA 468
DB 184 GCTCTCCCGGAGACCAAGACTGGAGGAGGATGGGATCGTGAGCCCGGTGAAGAGCCAG 243

QY 469 GGAGCTGTGATCTTGCTGACCTTTCAGTACTACTGAGCACTAGAGGCTGTGATATACA 528
DB 244 GGTCACTGTGTTCTGTGTGACCTTCAGCACCCTGTTCTTCTTGAGGAGCATATACT 303

QY 529 CAGCTAACTGGAAGAGCACAATATATCTGAAACAGCAACTTGTGTGAGCTGTGCTCAGCA 588
DB 304 CAGGCACTGGGAGGAGCTGCTCTCTTCTGAGCAGCAGCTGTTGATTTGTGCTACTGCA 363

QY 589 TTCATTAATTTGGATGCAATGGAGTTTGCCTTCCCAAGCCTTTGATAGTTAACTAC 648
DB 364 TACAATAATTTCCGATGCAATGGAGGCTTACCCTTCCAGGCTTTTGTGATCAATAATAC 423

QY 649 AATGAGGATCGACACAGACAGACTTATCCATACCTTGTGTCAATGGTATCTGCAAC 708
DB 424 AATGAGGCTTTGACACTGAAAGAGCTTACCCTTACAGGCTGTCAATGGCATCTGTCAT 483

QY 709 TTCAGCAGGAGATTTGTTGTCAGCTCATTTGATTCGATTAACATCACCTGGTGCT 768
DB 484 TACAAGCCTGAAACGTTGGAGTCAAGGTTTGGACTCCGTTAATCATCACCTGGTGCT 543

QY 769 GAGGATGAGTTGAAGCATGCAAGTGGCTTGGTGGCTCCAGTTAGCTTTCATTTGAGGTT 828
DB 544 GAGGATGAGCTGAAGAAATGCTGTTGGACTTGTGTCAGTTAGTTGCTTTCAGGTG 603

QY 829 GTGAAGGTTTCAATCTGTACAGAAAGTGTATACAGCAGTGAACACTGTGGGAAGAT 888
DB 604 ATCAACGCTTTTCAAGATGATCAAGAGTGGAGTTTACACAAGTGAACCTTGTGGAATCT 663

QY 889 CCAATGGATGTAACACGAGTCTTTCGCTGCTGATGAGTTCGAGGAGCGGATTCCT 948
DB 664 CCAATGGATGTAACACGCTGTTCTGCGGCTTGGCTATGTTGTGAAATGGCGTTCCC 723

QY 949 TATTGGCT 956
DB 724 TACTGGCT 731

```

RESULT 8

CB982399 748 bp mRNA linear EST 01-MAY-2003
LOCUS CAB70006 IVaF A04 Cabernet Sauvignon Berry Post-Veraison - CAB7
DEFINITION Vitis vinifera cDNA clone CAB70006 IVaF_A04 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

FEATURES
source

CB982399
CB982399.1 GI:30305605
EST.
Vitis vinifera
Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; Vitaceae; Vitis.
1 (bases 1 to 748)
Goes da Silva,F., Iandolo,A., Lim,H., Baek,J., Jones,K. and
Cook,D.
Expressed sequence tags from cabernet sauvignon berries at various
developmental stages
Unpublished (2003)
Contact: Douglas Cook, PhD
CABES Genome Facility
UC Davis, Plant Pathology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcook@ucdavis.edu
Seq primer: ACGTACCGACATATGCC.
Location/Qualifiers
1..748
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAB70006 IVaF_A04"
/sex="Hermaphrodite"
/dev_stage="Post-Veraison, 18-19 brx"
/lab_host="DH5alpha"
/notes="Organ: Berry; Vector: pDNR; Site 1: SfiI; Site 2:
SfiI; CAB7 is a cDNA library of Cabernet Sauvignon clone 8
berries. Samples were collected post-veraison from
field-grown vines during stage III of berry growth at 89
days after full bloom. Berries soluble solid concentration
ranged between 18-19 brx. Sampled vines were located at
the University of California, Davis, Experimental
Vineyard. cDNAs were made by oligo-dT priming and direct
ly cloned. 5' and 3' adaptors were used in cloning as
follows: 5'-AAGCAGTGTATCAACGAGTGGCCATTCAGCCGGG-3' and
5'-ATTCTAGAGCGGAGCGCCACATG-3' (30)NN-3'. Library was
constructed using the Clontech Creator SMART kit and
size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

```

Query Match      38.1%; Score 407.8; DB 14; Length 748;
Best Local Similarity 71.6%; Pred. No. 7.9e-88;
Matches 535; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 214 GAGATCAAGAAGGTTCCGGATCTTCGTGGAGAATCTAGCGTTTATCCGGTCCACTAAT 273
DB 2 GAGATTAAGCTGAGATTCGAGATTTTTCGGAGAATTTGAAATCATCATGATCCACCAAC 61

QY 274 CGAAGAGATCTGTCTGATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGAGGAA 333
DB 62 AGAAAGGGCTTGCCTTTATATCTAGCTGTATTAATCAGTTTCGCTGATTGGACCTGGGAAGAG 121

QY 334 TTCGCGACCAATCGCTTGTGTCGGCGCAGAACTGCTCGCGCACTCGCATGGAACCAAC 393
DB 122 TTCGCGAGACACAGGTTGGAGCTGTCAGAACTGCTCTGCCACCTTGAAGGCAATCAC 181

QY 394 CGGTTTGTGATGGCTGCTTCTGTGTAACGAGGATTTGAGGAGCAAGGATATGATGAGC 453
DB 182 AAGCTAACTGACGTTATCTCTCTGAGACGAAGATTGGAGAGAGATGGCATATATTCAGC 241

QY 454 CTGTAAAGGCCAAGAGAGCTGTGATCTTCTGACCTTTCAGTACTTTCAGTACTGAGCACTA 513
DB 242 CCAATCAAGATCAAGGTCAGTCTGTGGATCTTGTGCACTTTCAGCACTTTCAGGAGCTTA 301

QY 514 GAGGCTGCATATACACAGCTAACTGGAAGAGACCATCATATTCTGGAACAGCAACTTGTG 573

```

Db 302 GAGGACGCTAGCGTACGCTAGGCAATTTGGGAAGGGGATCTCTCTCTGAGCAGCAGCTTGTG 361
 QY 574 GACTGTGCTCAGCATTCAATTAATTTGATGCAATGAGGTTTGCCTTCCCAAGCCTTT 633
 Db 362 GACTGTGCGGAGCTTCAATTAATTTGATGCAATGAGGTTTGCCTTCCCAAGCCTTT 421
 QY 634 GAATAGCTTAAGTACAAATGAGGCAATCGACACAGAACAGATTAATCCATACCTTGGTGTG 693
 Db 422 GAGTACATCAATACAAATGAGGCAATCGACACAGAACAGATTAATCCATACCTTGGTGTG 481
 QY 694 ATGGTATCTGCACTTCAAGCAGGAGATTTGTTGTCAGGTCATGATTCGATTAAC 753
 Db 482 GATGGCACCCTGCAAAATTTCTTCAGAAATATTTGGTGTTCAGTCTCGACTCTGTGAAT 541
 QY 754 ATCACCTCTGGTCTCAGGATGAGTTGAAGCATGACAGTGGCTTGGTGGTCCAGTTAGC 813
 Db 542 ATTACCTCTGGTCTCAGGATGAGTTGAAGCATGACAGTGGCTTGGTGGTCCAGTTAGC 601
 QY 814 GTTGCAATTTGAGTTGAAAGTTTCAATCTGTACAGAAAGGTTATACAGCAGTGAC 873
 Db 602 GTGGCAATTTGAGTGTGCTCATGATTTCCGATTTTACAGAAAGGAGTTTACACAAAGTGA 661
 QY 874 ACCTGTGGAAGACATCAATGAGATGTGAACCAAGCAGTTCTTGCCTGCTTATGAGTGC 933
 Db 662 ACTTGTGGCAGCACTCCATGATGTGACCACTGCTTCTTGCAGTCTGGGTATGGAGTT 721
 QY 934 GAGGACGGGATCTCTTATTTGGCTCATC 960
 Db 722 GAAGATGGTGTAGCACTAGGCTCATC 748

RESULT 9
 CK268548
 LOCUS
 DEFINITION EST714626 potato abiotic stress cDNA library Solanum tuberosum cDNA clone POACH43 5' end, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Solanum tuberosum (potato)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiales; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 936)
 Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A. and Baker, B.
 Generation of ESTs from abiotic stressed potato tissue
 Other ESTs: EST714627
 Unpublished (2003)
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from TIGR via potato@tigr.org
 Seq primer: Art TAG GTG ACA CTA TAG.

FEATURES
 source
 1..936
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="POACH43"
 /tissue_type="abiotic stress treated leaf and root tissue"
 /lab_host="DH10B-TonA"
 /clone_lib="potato abiotic stress cDNA library"
 /note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI; supplier: Solanum tuberosum var. Kennebec plants were grown from cuttings on a 16hr light/8 hr dark cycle at 25 C for 3-4 weeks. Abiotic stress conditions were applied to four separate sets of plants. Set 1 involved saturation of the soil with 150 mM NaCl and tissues were harvested at following application of the salt stress (leaves: 2hr,

6hr, 12hr, 1d, 2d, and 4d; roots: 2hr, 6hr, 12hr, and 2d). Set 2 were grown under the standard conditions and then were water stressed by withdrawal of further watering applications. Drought stressed plants were harvested after cessation of watering (leaves: 3d, 5d, and 7d; roots: 3d and 5d). Set 3 were grown under the standard conditions and then were cold stressed by placement at 4 C. Cold stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d, and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d, then were heat stressed by placement at 35 C. Heat stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, 2d and 4d and heat-stressed roots were harvested at 6 hr, 12 hr, 1 d, and 4d. RNA was isolated from all tissues and equal RNA from each tissue and stress was pooled to construct the cDNA library. RNA sample."

ORIGIN

Query Match 37.9%; Score 405.6; DB 14; Length 936;
 Best Local Similarity 68.0%; Pred. No. 3e-87;
 Matches 562; Conservative 0; Mismatches 264; Indels 0; Gaps 0;
 QY 107 CTGACTGATTGAGCCTGCCATCTCGGCGTCTTGGCAGTTGCGCGCAGCCTTCCT 166
 Db 102 CCGAAGAGCTGGAGACGGAAATCTTCAAGTGTGCGGCCAGACTCGGAATGCTCTCTCT 161
 QY 167 TCGCACGCTTCCGCCGAGGTACCGGAAGAGCTACGGATCGGAGGAGGATCAAGAAGA 226
 Db 162 TCGTCTGCTTGTCTATCAGGCATCGGAAGAGTACGAGTCCGTTGAGGAGATCAAGCAA 221
 QY 227 GGTTCGGGATCTTCGTGGAGATCTAGCGTTTATCCGGTCCACTAATTCGAGGATCTGT 286
 Db 222 GGTTCGAGATATTTTGGACAACTCTGAAGATGATCCGATCGCATACGCAAGCAGGACTAT 281
 QY 287 CGTATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGGAATTCGGGACCAATC 346
 Db 282 CATCAAACTCGGTGTCAATGAGTTTACCGACCTAACATGGATGAGTTCCGTAGACACA 341
 QY 347 GCCTTGTGCGCGCAGAACTGCTCGGGAATCGGCTGGAACCAACCGGTTTTCGATG 406
 Db 342 AGTTGGGGCATCTCAAAACTGTTCTGCCACTACAAAGGGCAATCTCAAGTAACTAACG 401
 QY 407 CGGTGCTTCTGTAACGAGGATTTGGGGAGCAAGGATAGTAGGCCCTCTAAGGACCC 466
 Db 402 TCGTCTTCCAGAGACGAGGAGCTGGAGGAGATGTTATTGTTAGCCACTGAGGCAC 461
 QY 467 AAGGAAGCTGTGATCTTGTCTGGATTTTCACTACTATGAGGACCTAGAGGCTGCATATA 526
 Db 462 AGGGCAAGTGGGATCTTGTCTGGACATTCAGCACTACTGTGTCACCTAGAGCAGCATATG 521
 QY 527 CACAGCTAACTGGGAAGGACACATCATTTCTGTAACAGCACTTGTGACCTGTGCTCAG 586
 Db 522 CCCAAGCAATTTGGGAAGGGAATCTCTCTGTGAGAGCAGCAGCTTGTGACTGTGCTGGAG 581
 QY 587 CATTTCAATAAATTTGGATGCAATGAGGTTTGCCTTCCAAAGCCTTTGAAATACGTTAAGT 646
 Db 582 CTTTTTAATACTTTGGCTGCAATGGGGGTTGCCATCACAAAGCCTTTGAGTACATTAAT 641
 QY 647 ACAATGGAGGATCGACACAGACAGACTTATCCATACCTTGTGTCATGCTATCTGCA 706
 Db 642 TCAATGGTGGTCTTGACACTGAAGAAGCATATCCATACACCGCAAGAAATGGCATATGTA 701
 QY 707 ACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCATTGATTTCGATAAACATACCTTGGGTG 766
 Db 702 AATTTCTCAGCAATATTTGGTGTCAAGTCATCAGTTCTGTCAATATTACCTTGGGTG 761
 QY 767 CTGAGGATGAGTTGAAGCATGAGTGGGCTTGGTGGCTCCAGTTAGCCCTTGCATTTGAGG 826
 Db 762 CTGAAGATGAATGAATAACCGCAGTTGCAATTTGGTTAGGCCCTGTGTAGTTGCTTTTGAAG 821
 QY 827 TTGTGAAAGTTTCAATCTGTACAAAGAAAGGTGTATACAGCAGTGCACACCTGTGTGAAGAG 886
 Db 822 TGTATAAAGGTTTCAACAGTATATAGAGCGGAGTTTACACAGCACTGAATGTGGCGACA 881

/note="Organ: Bud; Vector: pDNR; Site 1: Sfil; Site 2: Sfil; CABUD is a cDNA library of Vitis vinifera cv. 'Cabernet Sauvignon' clone 8 dissected buds. Samples were collected May 13, 2002 from pre-bloom plants (10-11 days before bloom), pre-veraison. Sampled vines were located at the University of California, Davis, Experimental vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGGTATCAACGAGAGTGGCCATACGGCCGGG-3' and
 5'-ATTCTAGAGCCGAGCGCGACATG-dT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Query Match 37.6%; Score 402.8; DB 14; Length 848;
 Best Local Similarity 76.0%; Pred. No. 1.4e-86;
 Matches 497; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 412 CTTCTGTAAAGAGGATTCGAGGAGCAAGGATAGTGAGCCCTTAAGGACCAAGCA 471
 Db 2 CTTCTGTAAAGAGGATTCGAGGAGCAAGGATAGTGAGCCCTTAAGGATCAAGGT 61

Qy 472 AGCTGTGGATCTGCTGACTTTTCAGTACTTGGAGCACTAGAGGCTGCATATACACAG 531
 Db 62 CACTGTGGATCTGCTGACTTTTCAGCACCACTGGAGCTTAGAGGAGCTTACGCTCAG 121

Qy 532 CTAAGTGAAGAGCACATCATTTATCTGAACAGCAACTTGTGGACTGTGCTCAGCATTC 591
 Db 122 GCATTTGGGAAGGGATCTCTCTGTCTGAGCAGCAGCTTGTGACTGTGCGGAGGTTTC 181

Qy 592 AATACTTTTGATGCAATGAGGTTTCCCTTCCCAAGCCCTTCAATACGTTAGTACAAAT 651
 Db 182 AATACTTTTGGTCCCGGGGATTCCTATCCCAAGCTTTTGAAGTACATCAATACAT 241

Qy 652 GGAGGATCGACACAGACAGACTTATCCATACCTTGTGTCAATGTATCTGCAACTTC 711
 Db 242 GGTGGCCTTGATCTAGGAAGCATATCTTACACTGCACTAGATGCACTGCAAAATT 301

Qy 712 AAGCAGAGATGTGTGTGTCAGGTCATTTGATTCGATTAACATCACTGGGTGCTGAG 771
 Db 302 TCTTCGAAATATTTGTGTTCAGTTCTCGACTCTGTGAATATACCTTGGGTGCTGAA 361

Qy 772 GATGAGTTGAAGCATGCAAGTGGGCTTGGTGGCTCCAGTTAGCGTTGCATTTGAGTTGTG 831
 Db 362 GATGAGTTGAAGCATGCAAGTGGGCTTGGTGGCTCCAGTTAGCGTTGCATTTGAGTTGTG 421

Qy 832 AAGGTTTCAATCTGTACAGAAAGGTGTATACAGAGTACACCTGTGGAAGATCCA 891
 Db 422 CATGATTTCCGATTTTACAGAAAGGAGTTTACACAGTGGAACTTGTGGCAGCACTCCC 481

Qy 892 ATGATGTGAACACGAGTCTTTCGCGTGGTGTATGAGTTCGAGGACGGGATTCCTTAT 951
 Db 482 ATGATGTGAACACGAGTCTTTCGCGTGGTGTATGAGTTCGAGTTCGAGTTCGAGTTC 541

Qy 952 TGGCTCATCAAGAACTCATGGGGTACAAAATTGGGGTGCATAGGCTACTTTAGATGGAA 1011
 Db 542 TGGCTCATCAAGAACTCATGGGGTACAAAATTGGGGTGCATAGGCTACTTTCAAAACGGAG 601

Qy 1012 CTCGCAAGACATGTGGTGTTCGCAACTTTCGCACTTTCATCCATTTGGTGGCT 1065
 Db 602 TTGGCAAGAAATATGTGGTGTGCAACTTTCGCACTTTCATCATATCCCTTGTGGCT 655

RESULT 12
CF510584/cLOCUS
DEFINITION

CF510584 886 bp mRNA linear EST 09-SEP-2003
 CABUD0001_IR_F08 Vitis vinifera cv. cabernet sauvignon (Clone 8)
 Bud - CABUD Vitis vinifera cDNA clone Cabud0001_IR_F08 3', mRNA
 sequence.

ACCESSION
VERSION
KEYWORDS

CF510584.1 GI:34542352
 EST.

SOURCE
ORGANISM

Vitis vinifera
 Vitis vinifera
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; Vitaceae; Vitis.

REFERENCE

1 (bases 1 to 886)
 Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and
 Cook, D.

AUTHORS

Expressed sequence tags from Vitis vinifera 'Cabernet Sauvignon'
 berries at various developmental stages

TITLE

JOURNAL

COMMENT

Unpublished (2003)
 Contact: Douglas Cook, PhD
 CARS Genome Facility
 UC Davis, Plant Pathology
 One Shields Ave, Davis, CA 95616, USA
 Tel: 530 754 6561
 Fax: 530 754 6617
 Email: drcocook@ucdavis.edu
 Seq primer: GCCAAGCAATGGTCTAG.

FEATURES

source

1..886
 Location/Qualifiers
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultivar="Cabernet Sauvignon (Clone 8)"
 /db_xref="taxon:29760"
 /clone="Cabud0001_IR_F08"
 /sex="Hermaphrodite"
 /dev_stage="Pre-bloom (10-11 days before bloom)"
 /lab_host="DHSalpha"
 /clone_lib="Vitis vinifera cv. cabernet sauvignon (Clone 8) Bud - CABUD"
 /note="Organ: Bud; Vector: pDNR; Site 1: Sfil; Site 2: Sfil; CABUD is a cDNA library of Vitis vinifera cv. 'Cabernet Sauvignon' Clone 8 dissected buds. Samples were collected May 13, 2002 from pre-bloom plants (10-11 days before bloom), pre-veraison. Sampled vines were located at the University of California, Davis, Experimental vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGGTATCAACGAGAGTGGCCATACGGCCGGG-3' and
 5'-ATTCTAGAGCGCGCGACATG-dT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Query Match 37.5%; Score 401.8; DB 14; Length 886;
 Best Local Similarity 76.0%; Pred. No. 2.4e-86;
 Matches 496; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 413 TTCCTGTAAAGGAGGATTCGAGGAGCAAGGATAGTGAGCCCTTAAGGACCAAGGAA 472
 Db 886 TTCCTGTAGACGAAAGATTGGAGAGAGATGGCAGCTAGTCAAGCCCAATCAAGATCAAGTTC 827

Qy 473 GCTGTGGATCTTGTGGACTTTTCACTACTTCTGGAGCACTAGAGGCTGCATATACACAGC 532
 Db 826 ACTGTGGATCTTGTGGACTTTTCAAGCACTTGGAGCTTAGAGGAGCTTACGCTCAGG 767

Qy 533 TTAATGGAAAGACGACATCATTTCTGAACAGCACTTGTGGACTGTGCTCAGCATTC 592
 Db 766 CATTTGGGAAGGGGATCTCTCTGTCTGAGCAGCAGCTTGTGACTGTGCCGAGGCTTTCA 707

Qy 593 ATAACTTTGGATGCAATGAGGTTTTCCTTCCAGGCTTTTGAATACGTTTAAAGTACAATG 652
 Db 706 ATAACTTTGGTGGTCCAGGGGATTTGCCATCCCAAGCTTTTGGATACATCAATACAATG 647

Qy 653 GAGGCAATCGACACAGAAAGACTTATCATCTTGTGTCAATGGTATCTGCACTTCA 712
 Db 646 GTGGCTTGTATCTAGGAAGCATATCTTACCTGGAAGTATGAGTGGTGGTGGTGGTGGTGG 587

Qy 713 AGCAGGAGATGTTGTTGTGTCAGGTCATTGATTCGATAAATCAATCACTCCCTGGTGGTGG 772
 Db 586 CTTCAAGAAATATTTGGTGTTCAGTTCGACTCTGTGAATATTACCTCGGTGGTGGTGGTGA 527

QY 773 ATGAGTTCAAGCATCGACGTGGCTTGGTCGCTCCAGTTAGCGTTAGCATTTCAGGTTTGA 832
 DB |||||
 QY 526 ATGAGTTAAAGCATCGACGTTCGATTGTCTCCAGTGAGTGTGGCATTTTGGTGTGCC 467
 DB |||||
 QY 833 AAGGTTTCAATCTGTACAAAGAAAGTGATATACAGCAGTGACACCTGTGGAGAGATCCAA 892
 DB |||||
 QY 466 ATGATTTCCGATTTTACAAAGAAAGAGTTTACAAAGTGGAACTTTGTGGCAGCACTCCCA 407
 DB |||||
 QY 893 TGGATGTAAACCAAGCAGTCTTCCGCTCGGTTATGAGTGCAGACGGGATTCCTTATT 952
 DB |||||
 QY 406 TGGATGTAAACCAAGTCTTCTTCAGTCGGGTATGGAGTTGAAGATGGTGTAGCACTACT 347
 DB |||||
 QY 953 GGCCTCATCAAGAACTCATGGGTGACAAATGGGGTGACAAATGGCTACTTTTAAAGATGGAAC 1012
 DB |||||
 QY 346 GGCCTCATCAAGAACTCATGGGGAGAAACTGGGGGACAAATGGCTACTTCAAAACGGAGT 287
 DB |||||
 QY 1013 TCGGCAAGAACATGTGTGGTGTGCACTTGGGCACTTATCCCAATTGTGGCT 1065
 DB |||||
 QY 286 TGGGCAAGAAATGTGTGGTGTGGCACTTGTCTATCATATATCCGTTTGTGCT 234
 DB |||||

RESULT 13

CK295860 977 bp mRNA linear EST 15-DEC-2003
 LOCUS EST758574 Nicotiana benthamiana mixed tissue cDNA library,
 DEFINITION normalized, full-length Nicotiana benthamiana cDNA clone NBMCX76 5',
 end, mRNA sequence.

ACCESSION

CK295860

VERSION

CK295860.1 GI:39880669

KEYWORDS

EST.

SOURCE

Nicotiana benthamiana

ORGANISM

Nicotiana benthamiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Nicotiana.
 1 (bases 1 to 977)
 Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
 Staskiewicz, B., Jin, H. and Baker, B.
 Generation of EST sequences from Nicotiana benthamiana
 Unpublished (2003)
 Other ESTs: EST758575
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from TIGR via potato@tigr.org
 Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

Location/Qualifiers
 1..977
 /organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NBMCX76"
 /tissue_type="abiotic and biotic stress-treated leaves,
 callus tissue and root tissue"
 /lab_host="DH10B-TonA"
 /clone_lib="Nicotiana benthamiana mixed tissue cDNA
 library, normalized, full-length"
 /note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: RNA was isolated from Nicotiana benthamiana
 tissues that include callus, roots from liquid culture
 grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
 cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
 challenged leaves (Pseudomonas syringae pv tomato 12 hr;
 Xanthomonas campestris pv campestris 12 hr, 18hr;
 Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
 campestris pv vesicatoria 18hr). RNA was isolated from
 these tissues and pooled in approximately equal molar
 amounts."

ORIGIN

Query Match

37.2%; Score 398.6; DB 14; Length 977;

Best Local Similarity 65.3%; Pred. No. 1.5e-85;
 Matches 601; Conservative 0; Mismatches 319; Indels 1; Gaps 1;
 QY 13 CTCCTCTTCTTCTCTACTCTTCTCGTCTCCGCCCTCTCCGTCCGCTCCGCGTTTCGAA 72
 DB |||||
 QY 45 CTCGTCTCGCGGTGGCCCTTTTCGCGCGGCACATTGCCGACCGGCTACCTTCGCGGAT 104
 DB |||||
 QY 73 GAAGACAATCCATCCGGTCCGTTACACAAAGCCCTGACTCGATTGAGCCTGCCATCCTC 132
 DB |||||
 QY 105 GAGATCCCGATCAGACAAATTTGTTCTGACGGTTTACATGAGCTGGAGAACGGAATTC 164
 DB |||||
 QY 133 GCGCTCTTGGCGATTGCCGCCACGCTTCCACTTCGACGGTTCGCCCGCAGGTACGGG 192
 DB |||||
 QY 165 CAACTCGTGGCAAGACTCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 224
 DB |||||
 QY 193 AAGAGCTACGATCGGAGGAGATCAAGAGAGGTTCCGGATCTCTCGTGAGAACTCTA 252
 DB |||||
 QY 225 AAGAGTACGACACAGTTGAGGAGTAAAGCAAGTTTCGAGTATTTCTTGGCAATTTG 284
 DB |||||
 QY 253 GCGTTTATCCGCTCCACTAATCGGAAGGATCTGTGTATACCTAGGAATCAACCAATTC 312
 DB |||||
 QY 285 AAGATGATTGATCGCATTAACAGAAAGGACTATCATACAAATCTCGTGTCAATGAGTTT 344
 DB |||||
 QY 313 GCGCACTGACCTGGGAGGAATTCGGACCAATCGCTTGGTGGCGCGCAGAACTGCTCG 372
 DB |||||
 QY 345 ACCGACATAACATGGGACGAGTTCCGGAGAGACAGTTGGGAGAGCTCAAAACTGTTC 404
 DB |||||
 QY 373 GCACATGCGCATGAAACCAACCGGTTTGTGATGGCGTCTTCTGTAAACGAGGATTCG 432
 DB |||||
 QY 405 GCCACCAAAAGGCAATCTCAAACTCACTAAGTTTGTCTCTCGCGAGACGAAAGACTGG 464
 DB |||||
 QY 433 AGGAGCAAGGATAGTAGCCCTGTAAAGACCAAGGAAGCTGTGGATCTTGTCTGACT 492
 DB |||||
 QY 465 AGGGAAGCTGGGATTTGTAGCCGAGTCAAGAACCCAGGGAAGTGGCGGATCTTGTCTG 524
 DB |||||
 QY 493 TTCAGTACTACTGGAGCACTTAGAGGCTGCATATACACAGCTAACTGGAAAGACACATCA 552
 DB |||||
 QY 525 TTACGACACTAGTGTGCACTAGAGCAGCATATGGCCAAGCATTTGGGAAGGGAATTTCT 584
 DB |||||
 QY 553 TTATCTGAACAGCAACTGTGTGACTGTGCTCAGCATTCATTAACCTTGGATCGCAATGGA 612
 DB |||||
 QY 585 CTATCTGAACAGCAGCTTGTGCACTGTGCGAGCTTTTAATAACTTTGGCTGCAATGGT 644
 DB |||||
 QY 613 GGTTCGCTTCCCAAGCCTTTTGAATACGTTAAGTACAATGGAGGCATCGACACAGAACAG 672
 DB |||||
 QY 645 GGGCTCCCATCAAGCCCTTTGAGTATATTAATCCANTGGTGTCTTGACACTGAAGAA 704
 DB |||||
 QY 673 ACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGATTTGGTGTCT 732
 DB |||||
 QY 705 GCATATCCATACACCGCAAGATGGCTTATGTAAATTTCTCATCAGAAAAATGTTGGTGT 764
 DB |||||
 QY 733 AAGTCAATTGATTCGATAACATCACCTGGTGGTGGAGTGGATGAGTGGAGCATGCAGTG 792
 DB |||||
 QY 765 AAAAGTCATCGATTCTGTCAATATTACCCCTGGGTGCTGGAAGATGAACCTAAATACGG 824
 DB |||||
 QY 793 GCGCTGTGCTGCTCCAGTTAGCGTTGCAATTTGAGGTTGTGAAAGGTTTCAATC-TGTACAA 851
 DB |||||
 QY 825 GCATTGGTTAGGCCGCTTAGTATAGCTTTTGTAGTGTATAAAGTTTCAACAAATACAA 884
 DB |||||
 QY 852 GAAAGGTGTATACAGCAGTGACACCTGTGGAGAGATCCCAATGGATGTGNAACACGAGT 911
 DB |||||
 QY 885 GAGTGGTGTATTACACCAGCACCGAATGTGGGCATCTCCCATGATGTATAAACCATGCTGT 944
 DB |||||
 QY 912 TCTTGGCGTGGTATGGAGT 932
 DB |||||
 QY 945 TCTTGGTGGGTACGGTGT 965
 DB |||||

RESULT 14

BE131652

LOCUS

BE131652

DEFINITION

L48-1651T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours

NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1651,

BE131652 720 bp mRNA linear EST 20-FEB-2001
 L48-1651T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
 NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1651,

```

mRNA sequence.
ACCESSION BE131652
VERSION BE131652.1 GI:8579015
KEYWORDS EST.
SOURCE Mesembryanthemum crystallinum (common iceplant)
ORGANISM Mesembryanthemum crystallinum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE Caryophyllales; Alzooaceae; Mesembryanthemum.
JOURNAL 1 (bases 1 to 720)
COMMENT Cushman, J.C.
An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
Unpublished (1997)
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: 248-17 row: E column: 3
Seq primer: T3
High quality sequence stop: 350
POLYA=No.

FEATURES             Location/Qualifiers
     source            1..720
     mol_type="mRNA"   /organism="Mesembryanthemum crystallinum"
     db_xref="taxon:3544"
     clone="L48-1651"
     tissue_type="Leaf, 48 h 0.4M NaCl"
     dev_stage="Six week old"
     clone_lib="Ice plant Lambda Uni-Zap XR expression
     library, 48 hours NaCl treatment"
     note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
     EcoRI; Site_2: XhoI"

ORIGIN
Query Match          36.8%; Score 394.2; DB 10; Length 720;
Best Local Similarity 72.2%; Pred. No. 1.5e-84;
Matches 513; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 350 TTGTTGGCGGCGCAACCTGCTCGCGACTGCGCATGCGCATGCGCAACCGGTTTGTTCGATGGCG 409
DB 10 TGGGAGCTGCCCAAACTGCTCTGCCACTCTCAAGGGCAGTCACAGCTCTCTCATGCGAG 69

QY 410 TGCTTCCTGTAAAGGAGATTGGAGGAGCAAGGATAGTGAGCCCTGTAAAGGACCAAG 469
DB 70 ACCTTCCTGAACAAAGATTTGGAGAGCTGAGGCGCATAGTCAGCCCGGTCAAGATCAAC 129

QY 470 GAAGCTGTGGATCTTGTGTGAGCTTTCAGTACTATGAGGAGCACTAGAGGCTGCATATACAC 529
DB 130 ACCACTGTGGATCTCTGTGGACATTCAGTACACTGGAGCTTTGGAGGCGAGCTTATGCGC 189

QY 530 AGCTTAACGGAAGAGGACATCATTTATCTGAACAGCAACTGTGCACTGTGCTTCAGCAT 589
DB 190 AGGCATTTGGAGAGAACATCTTTTGTCTGAGCAACACTTGTGATTTGCTGTGTCAT 249

QY 590 TCAATAACTTTGGATGCAATGGAGTTTGCCTTCCAGCCCTTTGAAATAGCTTAAGTACA 649
DB 250 TTGATAATCATGGTTGCAATGGTGGTTGGCCATCCCAAGCTTTTCGAGTATGTCAAGTACA 309

QY 650 ATGGAGGATCGACACAGCAAGACTTATCCATCTTGGTGTCAATGGTATCTGCAACT 709
DB 310 ATGGAGGACTTGACACAGAGAGGCTTATCTTACACAGCAAGAGTGTGAATGTAAT 369

QY 710 TCAAGCAGGAGAAATGTTGGTGTCAAGGTCAATGATTAACATCAACCCCTGGGTGCTG 769
DB 370 TCTCAGCCCAAAATGTTGGTGTTCAGTCTCTGGACTCGGTAAACATCACCCCTCGAGCTG 429

```

```

770 AGGATGAGTTGAAGCATGCACTGGGCTTGGTGGTCCAGTTAGCTTGCATTTGAGGTTG 829
DB 430 AAGAGGAACTTAAGGATGCTGTGGCAATGGTCCGCTCTGTGAGTGTGCGCATTTGAGGTTG 489

QY 830 TGAAGGTTTCAATCTGTACAAGAAAGGTTGTATACAGCAGTGACACCTGTGGAAGAGATC 889
DB 490 TTATGATTTCCGGTTTACACGAAGAGGTTTACACAGCAGCAACACTGTGGCAGCAGCC 549

QY 890 CAATGGATGTAACACGCGAGTTCTTCCCGTCGGTTATGAGATCGAGGAGCGGATTCCTT 949
DB 550 CCATGGATGTGAACCATGCTGTTTCTAGCTGTGGATATGAGTGGAGATGGTATCCCAT 609

QY 950 ATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGTGACAATGGCTACTTTAAAGATGG 1009
DB 610 ACTGGCTTATCAAGAACTCTGGGGAGAGGATTTGGGTGACAATGGATCTTCAAGATGG 669

QY 1010 AACTCGGCAAGAACATGTGGTGTTCGCACTTGGCGCATCTTATCCCATGG 1060
DB 670 AGATGGGAAAAAACATGTGTGGTGTGCTACTTGTGGCGCATATCCAGTTG 720

RESULT 15
CK294622      965 bp      mRNA      linear      EST 15-DEC-2003
LOCUS        EST757336 Nicotiana benthamiana mixed tissue cDNA library,
DEFINITION   normalized, full-length Nicotiana benthamiana cDNA clone NEMC076 5'
end, mRNA sequence.
ACCESSION    CK294622
VERSION      CK294622.1 GI:39878196
KEYWORDS     EST.
SOURCE       Nicotiana benthamiana
ORGANISM     Nicotiana benthamiana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 965)
Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
Staskiewicz, B., Jin, H. and Baker, B.
Generation of EST sequences from Nicotiana benthamiana
Unpublished (2003)
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from TIGR via potato@tigr.org
Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES             Location/Qualifiers
     source            1..965
     organism="Nicotiana benthamiana"
     mol_type="mRNA"
     db_xref="taxon:4100"
     clone="NEMC076"
     tissue_type="abiotic and biotic stress-treated leaves,
     callus tissue and root tissue"
     lab_host="DH10B-Tona"
     clone_lib="Nicotiana benthamiana mixed tissue cDNA
     library, normalized, full-length"
     note="Vector: pCMVSPORT6.1; Site_1: EcoRI; Site_2: NotI;
     supplier: RNA was isolated from Nicotiana benthamiana
     tissues that include callus, roots from liquid culture
     grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
     cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
     challenged leaves (Pseudomonas syringae pv tomato 12 hr;
     Xanthomonas campestris pv campestris 12 hr, 18hr;
     Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
     campestris pv vesicatoria 18hr). RNA was isolated from
     these tissues and pooled in approximately equal molar
     amounts."

ORIGIN
Query Match          36.5%; Score 390.8; DB 14; Length 965;
Best Local Similarity 64.9%; Pred. No. 1.2e-83;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: March 29, 2004, 21:10:04 ; Search time 86 seconds
(without alignments)
1156.474 Million cell updates/sec

Title: US-10-087-714-2
Perfect score: 1865
Sequence: 1 MAALKLFFLLFVLSVAL.....ELGKMGCVATCASYPIVAV 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1865	100.0	352	7	Adc61377 Vanilla p
2	1323	70.9	363	7	Adc61399 Zea mays
3	1302.5	69.8	360	7	Adc61398 Zea mays
4	1300.5	69.7	378	4	Aab65775 Cysteine
5	1300.5	69.7	378	4	Aab65775 Cysteine
6	1286	69.0	360	7	Adc61397 Nicotiana
7	1279.5	68.6	358	3	Aag17823 Arabidops
8	1279.5	68.6	358	3	Aag47118 Arabidops
9	1272.5	68.2	360	7	Adc61395 Arabidops
10	1270	68.1	365	2	Aaw89557 Triticum
11	1265.5	67.3	358	3	Aag53786 Arabidops
12	1261.5	67.6	362	7	Adc61400 Oryza sat
13	1255.5	67.3	362	7	Adc61396 Hordeum v
14	1222	65.5	325	3	Aag47119 Arabidops
15	1222	65.5	325	3	Aag17824 Arabidops
16	1219	65.4	325	3	Aag53787 Arabidops
17	1148	61.6	357	7	Adc61399 Zea mays
18	1133	60.8	283	3	Aag47120 Arabidops
19	1133	60.8	283	3	Aag17825 Arabidops
20	1128	60.5	283	3	Aag53788 Arabidops
21	1010.5	54.2	251	7	Abm74089 DNA clone
22	976.5	52.4	255	3	Aag35363 Zea mays
23	939	50.3	218	3	Aag35365 Zea mays
24	939	50.3	226	3	Aag35364 Zea mays
25	856.5	45.9	277	3	Aag25078 Arabidops

26	856.5	45.9	304	3	AAG25077 Arabidops
27	845.5	45.3	277	7	ADB94916 Programme
28	792	42.8	244	3	AAG25079 Arabidops
29	796	42.7	333	7	ADB93368 Rat Prote
30	759	40.7	335	7	ADB83370 Human Pro
31	759	40.7	364	3	AAB58140 Lung canc
32	759	40.7	365	4	AAO13548 Human pol
33	759	40.7	365	4	AAO13563 Human pol
34	656.5	35.2	225	4	AAB65779 Cysteine
35	656.5	35.2	225	7	ADB94798 Programme
36	656.5	35.2	326	2	AAR57071 Fasciola
37	644.5	34.6	326	2	AAR57080 Fasciola
38	609	32.7	470	7	ADB94914 Programme
39	594.5	31.9	452	7	ADB94912 Programme
40	594	31.8	500	4	AAB65766 Cysteine
41	594	31.8	500	7	ADB94785 Programme
42	589	31.6	464	2	AAW19542 Soybean t
43	589	31.6	464	5	ABB81808 Soybean t
44	589	31.6	464	6	AAO19627 Soybean t
45	589	31.6	464	6	ADA21044 Soy prote

ALIGNMENTS

RESULT 1
ADDC61377
ID ADC61377 standard; protein; 352 AA.
XX AC ADC61377;
XX AC ADC61377;
DT 18-DEC-2003 (first entry)
XX DE Vanilla planifolia 4-hydroxybenzaldehyde synthase (4HBS).
XX KW Vanillin production; plant; vanillin biosynthesis;
XX KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme.
XX OS Vanilla planifolia.

Key Location/Qualifiers
FT Misc-difference 161..162
FT Misc-difference 179..180
FT Misc-difference 179..180
FT Misc-difference 181..182
FT Misc-difference 199..200
FT Misc-difference 228
FT Misc-difference 228
US2003070188-A1.
10-APR-2003.
28-FEB-2002; 2002US-00087714.
15-JUL-1997; 97US-0052604P.
15-JUL-1998; 98MO-USO14895.
25-MAY-2000; 2000US-00462576.
28-FEB-2001; 2001US-0272415P.
(HAVK/) HAVKIN-FRENKEL D.
(PODS/) PODSTOLSKI A.
(DIXO/) DIXON R A.
Havkin-Frenkel D, Podstolski A, Dixon RA;
WPI; 2003-743794/70.
N-PSDB; ADC61376.
Improving vanillin production in cultured Vanilla planifolia by

PT supplementing tissue (e.g., embryo) culture of *V. planifolia* with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 XX Claim 18; Fig 6; 49pp; English.
 XX The present invention relates to a method for improving vanillin
 CC production in cultured Vanillin planifolia, and in intact plants. The
 CC method involves genetically engineering *V. planifolia* to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC synthase (4HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured *V. planifolia*. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present sequence represents *V. planifolia* 4HBS.
 XX Sequence 352 AA;

Query Match 100.0%; Score 1865; DB 7; Length 352;
 Best Local Similarity 100.0%; Pred. No. 5.6e-176; Indels 0; Gaps 0;
 Matches 352; Conservative 0; Mismatches 0;
 QY 1 MAAKLFFLLVLSVALAGFEEDNPISVTORPDSIEPAIILGVLSGCRHAFHAFPA 60
 DB 1 MAAKLFFLLVLSVALAGFEEDNPISVTORPDSIEPAIILGVLSGCRHAFHAFPA 60
 QY 61 RYKSGYSGEERIKKRFQIFVENLAFIRSTNRKDLSTLGINQADLTWEEFTRNLGAA 120
 DB 61 RYKSGYSGEERIKKRFQIFVENLAFIRSTNRKDLSTLGINQADLTWEEFTRNLGAA 120
 QY 121 QNCSTAHGNHRFVDGVLPTVTRDMREOGIVSPVKDQSGSWTSTTGALAAAYTQTGS 180
 DB 121 QNCSTAHGNHRFVDGVLPTVTRDMREOGIVSPVKDQSGSWTSTTGALAAAYTQTGS 180
 QY 181 TLSEQQLVDCASAFNFCGGLPSQAFYVYKNGIDTEQTPYLVGMGICNFKQENVG 240
 DB 181 TLSEQQLVDCASAFNFCGGLPSQAFYVYKNGIDTEQTPYLVGMGICNFKQENVG 240
 QY 241 KVIDSINITLGADELKHAAGLVREPVSVAFVVKGFNLKKGVSSTTCGRDPMVNHAV 300
 DB 241 KVIDSINITLGADELKHAAGLVREPVSVAFVVKGFNLKKGVSSTTCGRDPMVNHAV 300
 QY 301 LAVGYGVEDGIPYWLKNSWGTNMGDNGYFKWELGKNCMGVATCASYP 352
 DB 301 LAVGYGVEDGIPYWLKNSWGTNMGDNGYFKWELGKNCMGVATCASYP 352

RESULT 2
 ADC61399
 ID ADC61399 standard; protein; 363 AA.
 XX
 AC ADC61399;
 XX
 DT 18-DEC-2003 (first entry).
 DE
 DE Zea mays cysteine protease #2.
 XX
 XX Vanillin production; plant; vanillin biosynthesis;
 KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.
 XX
 OS Zea mays.
 XX
 FN US2003070188-A1.
 XX
 PD 10-APR-2003.
 XX
 PF 28-FEB-2002; 2002US-00087714.
 XX
 PR 15-JUL-1997; 97US-0052604P.
 PR 15-JUL-1998; 98WO-US014895.
 PR 25-MAY-2000; 2000US-00462576.
 PR 28-FEB-2001; 2001US-0272415P.
 XX

PA (HAYK/) HAVKIN-FRENKEL D.
 PA (PODS/) PODSTOLSKI A.
 PA (DIXO/) DIXON R A.

XX Havkin-Frenkel D, Podstolski A, Dixon RA;
 XX WPI; 2003-743794/70.

XX Improving vanillin production in cultured Vanillin planifolia by
 PT supplementing tissue (e.g., embryo) culture of *V. planifolia* with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 XX Disclosure; Fig 6; 49pp; English.

XX The present invention relates to a method for improving vanillin
 CC production in cultured Vanillin planifolia, and in intact plants. The
 CC method involves genetically engineering *V. planifolia* to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC synthase (4HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured *V. planifolia*. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present sequence represents a plant cysteine protease
 CC used for comparison to the *V. planifolia* 4HBS protein of the invention.
 XX Sequence 363 AA;

Query Match 70.9%; Score 1323; DB 7; Length 363;
 Best Local Similarity 70.8%; Pred. No. 3.3e-122;
 Matches 257; Conservative 30; Mismatches 64; Indels 12; Gaps 6;
 QY 1 MAAKLFPFLPFL----VSALSVALAGFEEDNPISVTORPDS-IPAILGVLSGCRHAF 54
 DB 1 MAHRRIILLAAVAATAVSAAASCFDDSNPIRPTVTDAAASALESTVPAALGRTRDAL 60
 QY 55 HPAPFARYKSGYSGEERIKKRFQIFVENLAFIRSTNRKDLSTLGINQADLTWEEFRT 114
 DB 61 RFAPFAVRYKSGYSGEERIKKRFQIFSESLQVLRSTNRKGLSYRLGYNRFADMSWEEFRA 120
 QY 115 NRLGAQNCSTAHGNHRF--VDGVLPTVTRDMREOGIVSPVKDQSGCS-WTSTTGALF 171
 DB 121 TRLGAAQNCSTALTGNERHMAAVALPETKDWREDGIVSPVKQHGCGSWTSTTGALF 180
 QY 172 AAYTQLTGSG--TLSEQQLVDCASAFNFCG-GLPSQAFYVYKNGIDTEQTPYLVGM 228
 DB 181 AAYTQATGKPTLSLSEQQLVDCGFAFNFGNCGGLPSQAFYIYKNGLDTEESYPTQGVN 240
 QY 229 GICNFKQENVGVKVIDSINITLGADELKHAAGLVREPVSVAFVVKGFNLKKGVSSTT 288
 DB 241 GICNFKQENVGVKVIDSINITLGADELKHAAGLVREPVSVAFVVKGFNLKKGVSSTT 300
 QY 289 CGRDPMDVNHAVLAVGYGVEDGIPYWLKNSWGTNMGDNGYFKWELGKNCMGVATCASYP 348
 DB 301 CGTTPMDVNHAVLAVGYGVEDGIPYWLKNSWGTNMGDNGYFKWELGKNCMGVATCASYP 360
 QY 349 IVA 351
 DB 361 IVA 363

RESULT 3
 ADC61398
 ID ADC61398 standard; protein; 360 AA.
 XX
 AC ADC61398;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Zea mays cysteine protease #1.
 XX
 KW Vanillin production; plant; vanillin biosynthesis;
 KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.

XX Zea mays.
 XX US2003070188-A1.
 XX 10-APR-2003.
 XX 28-FEB-2002; 2002US-00087714.
 XX 15-JUL-1997; 97US-0052604P.
 XX 15-JUL-1998; 98WO-US014895.
 XX 25-MAY-2000; 2000US-00462576.
 XX 28-FEB-2001; 2001US-0272415P.
 XX (HAVK/) HAVKIN-FRENKEL D.
 XX (PODS/) PODSTOLSKI A.
 XX (DIXO/) DIXON R A.
 XX Havkin-Frenkel D, Podstolski A, Dixon RA;
 XX WPI; 2003-743794/70.
 XX Improving vanillin production in cultured Vanillin planifolia by
 XX supplementing tissue (e.g., embryo) culture of V. planifolia with malic
 XX acid or by subjecting the culture to heat or mechanical shear stress.
 XX Disclosure; Fig 6; 49pp; English.
 XX The present invention relates to a method for improving vanillin
 XX production in cultured Vanillin planifolia, and in intact plants. The
 XX method involves genetically engineering V. planifolia to overproduce
 XX enzymes associated with step(s) involved in vanillin biosynthesis in the
 XX plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 XX synthase (HBS) enzyme are also provided. The method is useful for
 XX improving vanillin production in cultured V. planifolia. The method
 XX results in the production of cultured cells or plants producing at least
 XX twice to 10 times more vanillin than cells or plants not produced by the
 XX new method. The present sequence represents a plant cysteine protease
 XX used for comparison to the V. planifolia 4HS protein of the invention.
 XX SQ Sequence 360 AA;
 Query Match 69.8%; Score 1302.5; DB 7; Length 360;
 Best Local Similarity 71.1%; Pred. No. 3.5e-120;
 Matches 256; Conservative 26; Mismatches 69; Indels 9; Gaps 6;
 QY 1 MAKKLFFLLVLSALVALA--GPEEDNPIRSVTQRPDS--TEPAILGVLSGRHAFHA 57
 DB 1 MAPRRLLVLAVALAATAAANSGFADSNPIRPVTDRAASALESTVFAALGTRDALRFA 60
 QY 58 RFARRYKSGYSEEEIKKRFIFVENLAFIRSTNRKDLSTYLGINQFADLTWEETNRL 117
 DB 61 RFAVRKGSYESAAEVHKFRIFSESLQLVRSNKGSLYRLGINRFADMSWEEFRATRL 120
 QY 118 GAAQNCATAGNHRF--VDGVLVPTVRDWRBQGIYSPVKDQSCGS--WTFSTTGALBAAY 174
 DB 121 GAAQNCATLTGNHRRRAAVALPETKDRBQGIYSPVKDQSCGS--WTFSTTGALBAAY 180
 QY 175 TQLTGS--TLSEQQLVDCASAFNFC--GGLPSQAFYKYNKGIDTQTYPLGMVIGC 231
 DB 181 TQATGPISLSEQQLVDCGLAFNFCNGGLPSQAFYKYNKGIDTQTYPLGMVIGC 240
 QY 232 NFKQENVGVKIDSINITLGADELKHAFLVPRVSVAFVYKYNKGIDTQTYPLGMVIGC 291
 DB 241 KFNENNVGVKIDSVNITLGADELKHAFLVPRVSVAFVYKYNKGIDTQTYPLGMVIGC 300
 QY 292 DPMNVNHAFLVGYGVEDGIPYWLKNSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 351
 DB 301 TPMNVNHAFLVGYGVEDGIPYWLKNSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 360
 RESULT 4
 AAB65775

AAB65775 standard; protein; 378 AA.
 AAB65775;
 27-MAR-2001 (first entry)
 Cysteine protease #16.
 Cell death modulator; programmed cell death; PCD; apoptosis;
 forestry plant.
 Eucalyptus grandis.
 WO2000075331-A1.
 14-DEC-2000.
 02-JUN-2000; 2000WO-NZ000086.
 04-JUN-1999; 99US-00325932.
 (GENE-) GENESIS RES & DEV CORP LTD.
 (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 Flinn B, Lasham A;
 WPI; 2001-061724/07.
 N-FSDB; AAF44802.
 Novel defender against cell death polynucleotide useful for modulating
 programmed cell death pathway and specific development pathways in
 forestry plant.
 Claim 22; Page 112-113; 142pp; English.
 The present invention relates to coding sequences (see AAF44740-F44840
 and AAF44843-F44844) and proteins (see AAB657714-B65814) involved in
 programmed cell death (PCD; apoptosis). The coding sequences and proteins
 of the present invention are useful for modulating a PCD or cell death
 pathway and various developmental pathways in a forestry plant, by stably
 incorporating one of the present coding sequences into the genome of the
 forestry plant, where the coding sequence provides a PCD pathway that is
 not present in a native form of the forestry plant
 SQ Sequence 378 AA;
 Query Match 69.7%; Score 1300.5; DB 4; Length 378;
 Best Local Similarity 70.4%; Pred. No. 6e-120;
 Matches 247; Conservative 33; Mismatches 60; Indels 11; Gaps 5;
 QY 9 LLFLVLSALVALAGFEEDNPIRSVTQRPD----SIEPAILGVLSGRHAFHAFARFARYG 64
 DB 31 LLLVAVVVSAAASSFEESNFIPLP---PDGGLRDLESSIVQIVGTRHAPSFARFANRYG 87
 QY 65 KSYGSEBEIKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEETNRLGAAQNC 124
 DB 88 KRYETABEIKLRFEIFRENKLRIRSTNKGSLPTLGVNKFADMSWEEFRHRLGAAQNC 147
 QY 125 ATAGNHRFVDGVLVPTVRDWRBQGIYSPVKDQSCGS--WTFSTTGALBAAYTQLTGS--T 181
 DB 148 ATTGNKHLTDEALPEMKOWREKGIYSPVKDQSCGS--WTFSTTGALBAAYHQAQKQIS 207
 QY 182 LSEQQLVDCASAFNFC--GGLPSQAFYKYNKGIDTQTYPLGMVIGCINFKQENVGV 240
 DB 208 LSEQQLVDCAGAFNFC--GGLPSQAFYKYNKGIDTQTYPLGMVIGCINFKQENVGV 267
 QY 241 KVIDSINITLGADELKHAFLVPRVSVAFVYKYNKGIDTQTYPLGMVIGCINFKQENVGV 300
 DB 268 QVLDSVNITLGADELKHAFLVPRVSVAFVYKYNKGIDTQTYPLGMVIGCINFKQENVGV 327
 QY 301 LAVGYGVEDGIPYWLKNSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 351
 DB 328 LAVGYGVEDGIPYWLKNSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 378


```
QY      61 RYKSGYSEEBEIKRFGIFVENLAFIRSTNRKDLSTYTLGINOFADLTWBEFRTNRLGAA 120
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      66 HRYKREYSEVEIKQRFVFLNKLXIRSHRNKKGLSYKLVNEFTDLTWDFRRDLGAA 125
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY     121 QNCSTAHGNHRFVGVLVPTDRWREOGIVSPVKDQSGCS-WTFSTTGALAEAAVTLTG 179
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db     126 QNCSATTKNLKVTNVLPETKDWREAGIVSPVKQKGGCSWTFTTTGALEAAVYSAFG 185
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY     180 S--TLSEQQLVDCASAFNNFGC--GGLPQOAFYVVKYKNGGIDTEQTPYLVGMGICNFQOE 236
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db     186 KGISLSEQQLVDCASAFNNFGCNGGLPSQAFYIYKNSGLDTEEAIPYTGKNGLCKFSE 245
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY     237 NVGVKVIDSINITLGADELKHAVALRVSVAFVVKGFNLKKGVTYSSDTCCGRDPMDV 296
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db     246 NVGVKVIDSINITLGADELKHAVALRVSVAFVVKGFNLKKGVTYSSDTCCGRDPMDV 305
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY     297 NHAVLAVGVGVEDGTPYVNLKNSWGTNCGDNGYFQWELGKNMCGVATCASYPIVA 351
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db     306 NHAVLAVGVGVEDGTPYVNLKNSWGTNCGDNGYFQWELGKNMCGVATCASYPIVA 360
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
AAG17823
ID  AAG17823 standard; protein; 358 AA.
XX
AC  AAG17823;
XX
DT  17-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 18990.
XX
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-00301439.
XX
PR  25-FEB-1999; 99US-0121825P.
PR  05-MAR-1999; 99US-0123180P.
PR  09-MAR-1999; 99US-0123548P.
PR  23-MAR-1999; 99US-0125788P.
PR  25-MAR-1999; 99US-0126264P.
PR  29-MAR-1999; 99US-0126785P.
PR  01-APR-1999; 99US-0127462P.
PR  06-APR-1999; 99US-0128234P.
PR  08-APR-1999; 99US-0128714P.
PR  16-APR-1999; 99US-0129845P.
PR  19-APR-1999; 99US-0130077P.
PR  21-APR-1999; 99US-0130449P.
PR  23-APR-1999; 99US-0130510P.
PR  28-APR-1999; 99US-0130891P.
PR  30-APR-1999; 99US-0131449P.
PR  30-APR-1999; 99US-0132048P.
PR  04-MAY-1999; 99US-0132407P.
PR  05-MAY-1999; 99US-0132484P.
PR  05-MAY-1999; 99US-0132485P.
PR  06-MAY-1999; 99US-0132486P.
PR  06-MAY-1999; 99US-0132487P.
PR  07-MAY-1999; 99US-0132863P.
PR  11-MAY-1999; 99US-0134256P.
PR  14-MAY-1999; 99US-0134218P.
PR  14-MAY-1999; 99US-0134219P.
PR  14-MAY-1999; 99US-0134221P.
PR  14-MAY-1999; 99US-0134221P.
PR  18-MAY-1999; 99US-0134370P.
PR  18-MAY-1999; 99US-0134768P.
PR  19-MAY-1999; 99US-0134941P.
PR  20-MAY-1999; 99US-0135124P.
PR  21-MAY-1999; 99US-0135353P.
PR  24-MAY-1999; 99US-0135629P.
PR  25-MAY-1999; 99US-0136021P.
PR  27-MAY-1999; 99US-0136392P.
PR  28-MAY-1999; 99US-0136782P.
PR  01-JUN-1999; 99US-0137222P.
PR  03-JUN-1999; 99US-0137528P.
PR  04-JUN-1999; 99US-0137502P.
PR  07-JUN-1999; 99US-0137724P.
PR  08-JUN-1999; 99US-0138094P.
PR  10-JUN-1999; 99US-0138540P.
PR  10-JUN-1999; 99US-0138847P.
PR  14-JUN-1999; 99US-0139119P.
PR  16-JUN-1999; 99US-0139452P.
PR  16-JUN-1999; 99US-0139453P.
PR  17-JUN-1999; 99US-0139492P.
PR  18-JUN-1999; 99US-0139454P.
PR  18-JUN-1999; 99US-0139455P.
PR  18-JUN-1999; 99US-0139456P.
PR  18-JUN-1999; 99US-0139457P.
PR  18-JUN-1999; 99US-0139458P.
PR  18-JUN-1999; 99US-0139459P.
PR  18-JUN-1999; 99US-0139460P.
PR  18-JUN-1999; 99US-0139461P.
PR  18-JUN-1999; 99US-0139462P.
PR  18-JUN-1999; 99US-0139463P.
PR  18-JUN-1999; 99US-0139750P.
PR  18-JUN-1999; 99US-0139763P.
PR  21-JUN-1999; 99US-0139817P.
PR  22-JUN-1999; 99US-0139899P.
PR  23-JUN-1999; 99US-0140353P.
PR  23-JUN-1999; 99US-0140354P.
PR  24-JUN-1999; 99US-0140695P.
PR  28-JUN-1999; 99US-0140823P.
PR  29-JUN-1999; 99US-0140991P.
PR  30-JUN-1999; 99US-0141287P.
PR  01-JUL-1999; 99US-0141842P.
PR  01-JUL-1999; 99US-0142154P.
PR  02-JUL-1999; 99US-0142055P.
PR  06-JUL-1999; 99US-0142390P.
PR  08-JUL-1999; 99US-0142803P.
PR  09-JUL-1999; 99US-0142920P.
PR  12-JUL-1999; 99US-0142977P.
PR  13-JUL-1999; 99US-0143542P.
PR  14-JUL-1999; 99US-0143624P.
PR  15-JUL-1999; 99US-0144005P.
PR  16-JUL-1999; 99US-0144085P.
PR  16-JUL-1999; 99US-0144086P.
PR  19-JUL-1999; 99US-0144325P.
PR  19-JUL-1999; 99US-0144331P.
PR  19-JUL-1999; 99US-0144332P.
PR  19-JUL-1999; 99US-0144333P.
PR  19-JUL-1999; 99US-0144334P.
PR  19-JUL-1999; 99US-0144335P.
PR  20-JUL-1999; 99US-0144352P.
PR  20-JUL-1999; 99US-0144632P.
PR  20-JUL-1999; 99US-0144633P.
PR  21-JUL-1999; 99US-0144814P.
PR  21-JUL-1999; 99US-0145086P.
PR  21-JUL-1999; 99US-0145088P.
PR  22-JUL-1999; 99US-0145085P.
PR  22-JUL-1999; 99US-0145087P.
PR  22-JUL-1999; 99US-0145089P.
PR  22-JUL-1999; 99US-0145192P.
PR  23-JUL-1999; 99US-0145145P.
PR  23-JUL-1999; 99US-0145218P.
PR  23-JUL-1999; 99US-0145224P.
PR  26-JUL-1999; 99US-0145276P.
PR  27-JUL-1999; 99US-0145913P.
PR  27-JUL-1999; 99US-0145918P.
PR  27-JUL-1999; 99US-0145919P.
PR  28-JUL-1999; 99US-0145951P.
```

```

PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148684P.
PR 13-AUG-1999; 99US-0149336P.
PR 16-AUG-1999; 99US-0149317P.
PR 17-AUG-1999; 99US-0149426P.
PR 18-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150386P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151085P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151303P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156559P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159328P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.6%; Score 1279.5; DB 3; Length 358;
Best Local Similarity 67.0%; Pred. No. 6.7e-118; Indels 7; Gaps 4;
Matches 240; Conservative 39; Mismatches 72;

QY 1 MAAKLLF---FILFLVSALSVALAGFEEDNPFRSQTQRPDSIEPAILGLVLSGRHAFHFA 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MSAKTILSSVVLVWLVAASAAANIGFDSNPFRWVSDGLREVRSVSQILGQSRHVL SFA 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 58 RFARRYKSYSEBEIKRGRFIFVENLAFIRSTNEKDLSTLGNQFADLWEEPRTRL 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 RFTHRYGKYQNVEMKLRFSIFKENLDLIRSTNKKGLSVKLVGNQFADLWEEPRTKL 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 118 GAAQNCSTAHGNHRFVDGLPVTRDMREOGIVSPVKDQSGCS-WTFSTTGALEAAVYQ 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 121 GAAQNCSTALKGSHKVTEALPETKDWREDGLVSPVKDQSGCSWTFSTTGALEAAVHQ 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 177 LTGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVVKXNGGIDTEQTPYLVGMGICNF 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 181 AFGKGISLSEQQLVDCAGAFNNYGCNGGLPSQAFYIKSNGGLDTEKAYPTGKDETCF 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 234 KOENVGVKVIDSINITLGADELKHAVLVRPVSVAFEVWVGFMNLYKKGVYSSDTCGRDP 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 241 SAENVGVQVLSNVNITLGADELKHAVLVRPVSIAFEVIHSFRLYKSGVYTDSCGSTP 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 294 MDVNHAVLAAGYGVEDGDPYMLIKNSWGTNMGDNGYFKMELKKNMCGVATCASYIVA 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 301 MDVNHAVLAAGYGVEDGDPYMLIKNSWGDWGDGKGYFKMGMKKNMCGIATCASYIPVVA 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
AAG47118
ID AAG47118 standard; protein; 358 AA.
XX
XX AAG47118;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 59353.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; Genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 05-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 08-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
```

21-JUL-1999	PR
22-JUL-1999	PR
23-JUL-1999	PR
24-JUL-1999	PR
25-JUL-1999	PR
26-JUL-1999	PR
27-JUL-1999	PR
28-JUL-1999	PR
29-JUL-1999	PR
30-JUL-1999	PR
31-JUL-1999	PR
01-AUG-1999	PR
02-AUG-1999	PR
03-AUG-1999	PR
04-AUG-1999	PR
05-AUG-1999	PR
06-AUG-1999	PR
07-AUG-1999	PR
08-AUG-1999	PR
09-AUG-1999	PR
10-AUG-1999	PR
11-AUG-1999	PR
12-AUG-1999	PR
13-AUG-1999	PR
14-AUG-1999	PR
15-AUG-1999	PR
16-AUG-1999	PR
17-AUG-1999	PR
18-AUG-1999	PR
19-AUG-1999	PR
20-AUG-1999	PR
21-AUG-1999	PR
22-AUG-1999	PR
23-AUG-1999	PR
24-AUG-1999	PR
25-AUG-1999	PR
26-AUG-1999	PR
27-AUG-1999	PR
28-AUG-1999	PR
29-AUG-1999	PR
30-AUG-1999	PR
31-AUG-1999	PR
01-SEP-1999	PR
02-SEP-1999	PR
03-SEP-1999	PR
04-SEP-1999	PR
05-SEP-1999	PR
06-SEP-1999	PR
07-SEP-1999	PR
08-SEP-1999	PR
09-SEP-1999	PR
10-SEP-1999	PR
11-SEP-1999	PR
12-SEP-1999	PR
13-SEP-1999	PR
14-SEP-1999	PR
15-SEP-1999	PR
16-SEP-1999	PR
17-SEP-1999	PR
18-SEP-1999	PR
19-SEP-1999	PR
20-SEP-1999	PR
21-SEP-1999	PR
22-SEP-1999	PR
23-SEP-1999	PR
24-SEP-1999	PR
25-SEP-1999	PR
26-SEP-1999	PR
27-SEP-1999	PR
28-SEP-1999	PR
29-SEP-1999	PR
30-SEP-1999	PR
01-OCT-1999	PR
02-OCT-1999	PR
03-OCT-1999	PR
04-OCT-1999	PR
05-OCT-1999	PR
06-OCT-1999	PR
07-OCT-1999	PR
08-OCT-1999	PR
09-OCT-1999	PR
10-OCT-1999	PR
11-OCT-1999	PR
12-OCT-1999	PR
13-OCT-1999	PR
14-OCT-1999	PR
15-OCT-1999	PR
16-OCT-1999	PR
17-OCT-1999	PR
18-OCT-1999	PR
19-OCT-1999	PR
20-OCT-1999	PR

PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144086P.
PR	16-JUL-1999;	99US-0144085P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145132P.
PR	22-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147935P.
PR	09-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149923P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151068P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161408P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 67.9%; Score 1265.5; DB 3; Length 358;		
Best Local Similarity 66.5%; Pred. No. 1.6e-116;		
Matches 238; Conservative 39; Mismatches 74; Indels 7; Gaps 4;		
QY	1	MAAKLLF---FLFLVVSALSVALAGFEEDNPIRSVTQRPDSEPAILGVLSGRHAFHPA 57
Db	1	MSVKLNLSSILLILFAAASKEIGFDESNPINQWSDNLHELEDITVQILGSRHVSFS 60
QY	58	RFARRYKSYGSEEEIKKGFIVENLAFIRSTNRKDSYTLGINQFADLTWEEFRTNRL 117
Db	61	RFTHRYKKYQSVSEMKLRFVFKENLDLIRSTNKKGLSYKLSLNQFADLTWQEFQRYKL 120
QY	118	GAQNCSATAGNHRFPVDGLVPTWRBEOGIVSPVKDGGSGS-WTPSTTGALAAATQ 176
Db	121	GAQNCSATLKGSHKITEATVPTDKWREDGLIVSPVKECHGSCWTFSTTGALAAATQ 180
QY	177	LTGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYKYNKGIDTEQTPYVGVWGICNF 233
Db	181	AFCKGISLSEQLVDCAGTFNNFGCHGLPSQAFYKYNKGIDTEATPYTGKGGCKF 240
QY	234	KQENVGVKVIDSINITLGADELKHAAGLVPRVSAFVVKGNLYKGVGSSDTCGRDP 293
Db	241	SAKNIGVQRDSVNITLGADELKHAAGLVPRVSAFVVKGNLYKGVGSSDTCGRDP 300
QY	294	MDVNHAVLAGYGVDEGIPYWLITKNSWGTNWDNGYFKMELGKMGCVATCASYPIVA 351
Db	301	MDVNHAVLAGYGVDEDDVPYWLITKNSWGTNWDNGYFKMELGKMGCVATCASYPIVA 358
RESULT 12		
ADC61400		
ID ADC61400 standard; protein; 362 AA.		
XX		

AC ADC61400;
 XX 18-DEC-2003 (first entry)
 DT Oryza sativa cysteine protease.
 XX Vanillin production; plant; vanillin biosynthesis;
 KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.
 XX Oryza sativa.
 OS
 XX US2003070188-A1.
 PN 10-APR-2003.
 XX 28-FEB-2002; 2002US-00087714.
 XX 15-JUL-1997; 97US-0052604P.
 PR 15-JUL-1998; 98WO-US014895.
 PR 25-MAY-2000; 2000US-00462576.
 PR 28-FEB-2001; 2001US-0272415P.
 XX (HAVK/) HAVKIN-FRENKEL D.
 PA (PODS/) PODSTOLSKI A.
 PA (DIXO/) DIXON R A.
 XX Havkin-Frenkel D, Podstolski A, Dixon RA;
 PI WPI; 2003-743794/70.
 XX Improving vanillin production in cultured Vanillin planifolia by
 PT supplementing tissue (e.g., embryo) culture of V. planifolia with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 XX Disclosure; Fig 6; 49pp; English.
 XX The present invention relates to a method for improving vanillin
 CC production in cultured Vanillin planifolia, and in intact plants. The
 CC method involves genetically engineering V. planifolia to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC synthase (4HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured V. planifolia. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present sequence represents a plant cysteine protease
 CC used for comparison to the V. planifolia 4HBS protein of the invention.
 XX Sequence 362 AA;
 SQ
 Query Match 67.6%; Score 1261.5; DB 7; Length 362;
 Best Local Similarity 66.3%; Pred. No. 4.1e-116;
 Matches 240; Conservative 44; Mismatches 67; Indels 11; Gaps 6;
 QY 1 MAAKLFFLLFL-----VSALSVALAGFEDNPISRTORPDS-IEPAILVGLGSCRHAF 54
 DB 1 MAHRRIILLAAVAATAVAASAAAGSDFDNPISRTVDHAASALESTVIAALGRTRGAL 60
 QY 55 HFAEFARRYKSYGSEEEIKKPGIEVENLAFIRSTNRKDLSTVLGINQFADLTWEET 114
 DB 61 RFARFAVRGHRKRDGAEEVQRRIFSELSLELVRSNRRGLPRLGINFADMSWEEFQA 120
 QY 115 NRLGAQNCSSATAGNHRFVDG-VLEPVRDREQGIVSFVQDQSGCS-WTFSTTGALFA 172
 DB 121 SRLGAQNCSSATLGNHRMEDAPALPETKDREDGIVSPVQDQHCSCWPESTTGSLEA 180
 QY 173 AYTLQITGS--TLSEQQLVDCASAFNPGC-GGLPSQAFVYKXNGIDTEQYVYLGVNG 229
 DB 181 RYTQATGPPVSLSEQLADCATYNNFGSGGLPSQAFYIKYNGGLDTEEAPYPTGVNG 240
 QY 230 ICNFKDENVGKVIDSINTLTGAEDELKHAAGLVPRVSVAFVYKGFNLYKGVYSSDTC 289
 DB 241 ICHYKENAGHKVLDVSNITLVAEDELKNAAGLVPRVSVAFQVINGFRMYKSGVYSDHC 300

290 GRDPMVDNHAFLAVGYGVEDGIPYMLIKNSWGTWMDNGYFVKVBLGKNKCGVATCASYPI 349
 301 GTSFMDNHAFLAVGYGVENGVPYMLIKNSWGDWGDNGYFTWEMGKNKCGIATCASYPI 360
 350 VA 351
 361 VA 362

RESULT 13
 ADC61396
 ID ADC61396 standard; protein; 362 AA.
 XX
 AC ADC61396;
 XX 18-DEC-2003 (first entry)
 DT Hordeum vulgare cysteine protease.
 XX Vanillin production; plant; vanillin biosynthesis;
 KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.
 XX Hordeum vulgare.
 OS
 XX US2003070188-A1.
 PN 10-APR-2003.
 XX 28-FEB-2002; 2002US-00087714.
 XX 15-JUL-1997; 97US-0052604P.
 PR 15-JUL-1998; 98WO-US014895.
 PR 25-MAY-2000; 2000US-00462576.
 PR 28-FEB-2001; 2001US-0272415P.
 XX (HAVK/) HAVKIN-FRENKEL D.
 PA (PODS/) PODSTOLSKI A.
 PA (DIXO/) DIXON R A.
 XX Havkin-Frenkel D, Podstolski A, Dixon RA;
 PI WPI; 2003-743794/70.
 XX Improving vanillin production in cultured Vanillin planifolia by
 PT supplementing tissue (e.g., embryo) culture of V. planifolia with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 XX Disclosure; Fig 6; 49pp; English.
 XX The present invention relates to a method for improving vanillin
 CC production in cultured Vanillin planifolia, and in intact plants. The
 CC method involves genetically engineering V. planifolia to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC synthase (4HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured V. planifolia. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present sequence represents a plant cysteine protease
 CC used for comparison to the V. planifolia 4HBS protein of the invention.
 XX Sequence 362 AA;
 SQ
 Query Match 67.3%; Score 1255.5; DB 7; Length 362;
 Best Local Similarity 67.7%; Pred. No. 1.1e-115;
 Matches 243; Conservative 42; Mismatches 63; Indels 11; Gaps 8;
 QY 3 AKLFFFLFLVLSALVALA---GPEEDNPISRTORPDS-IEPAILVGLGSCRHAFHAF 58
 DB 4 ARVLLALAVLATAAVALVASSSSPADSNPIRPTVDRRAASTLESVALGALGTRIALAFAR 63
 QY 59 FARRYGKSYGSEEEIKKRFQIFVENLAFIRSTNRKDLSTVLGINQFADLTWEETNRLLG 118


```
Db 64 FAVRYGKSYSAEAEVRRRIIFSESLEEVSTNRKGLPYRLGICINRFSMSWEERFOATRLG 123
Qy 119 AAQNCSATAGNHRFVD-GVLPVTRDREOGIVSPVKDQSGS-WTPESTTGAAEAAYTO 176
Db 124 AACTCSAT-KGNHLEDAALPETKOWREDGIVSPVKQAHCGSWTFSTTGAAEAAYTO 182
Qy 177 LTGS---TLSEQLVDCASAFNFGC-GGLPSQAFYKYNKGIDTEQTPYPLGVNGICNF 233
Db 183 ATGKNISLSEQLVDCAGFNFGCGGLPSQAFYIKYNGIDTEESYPYKGVNGVCHY 242
Qy 234 KQNVGVKVIDSINITLGADELKNAVLVRVSV-AFEVVKGENLYKGVYSSDTCGRD 292
Db 243 KAENRAVOVLDSVNIINAEDELKNAVLVRVSVAAFOVIDGFRQYKSGVYTSDHGCTT 302
Qy 293 PWDVNAVLAVGVVEDGIPYMLIKNSGNTNMGDNGYFKMELGRNMGCVATCASYPIVA 351
Db 303 PDVNAVLAVGVGVGVYVYVLIKNMGADMGDNGYFKMEMGKNMCAIATCASYPIVA 361

RESULT 14
AAG47119
ID AAG47119 standard; protein; 325 AA.
XX AC AAG47119;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59354.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX KN termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 09-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0128264P.
XX PR 29-MAR-1999; 99US-0128785P.
XX PR 01-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
XX PR 08-APR-1999; 99US-0128714P.
XX PR 16-APR-1999; 99US-0129845P.
XX PR 19-APR-1999; 99US-0130077P.
XX PR 21-APR-1999; 99US-0130449P.
XX PR 23-APR-1999; 99US-0130510P.
XX PR 28-APR-1999; 99US-0130891P.
XX PR 30-APR-1999; 99US-0131449P.
XX PR 04-MAY-1999; 99US-0132407P.
XX PR 05-MAY-1999; 99US-0132485P.
XX PR 06-MAY-1999; 99US-0132486P.
XX PR 07-MAY-1999; 99US-0132487P.
XX PR 11-MAY-1999; 99US-0132863P.
XX PR 14-MAY-1999; 99US-0134256P.
XX PR 14-MAY-1999; 99US-0134218P.
XX PR 14-MAY-1999; 99US-0134219P.
XX PR 18-MAY-1999; 99US-0134370P.
XX PR 19-MAY-1999; 99US-0134768P.
XX PR 20-MAY-1999; 99US-0134941P.
XX PR 21-MAY-1999; 99US-0135124P.
XX PR 25-FEB-1999; 99US-0135353P.
XX PR 25-MAY-1999; 99US-0135629P.
XX PR 27-MAY-1999; 99US-0136021P.
XX PR 28-MAY-1999; 99US-0136392P.
XX PR 01-JUN-1999; 99US-0136782P.
XX PR 03-JUN-1999; 99US-0137222P.
XX PR 04-JUN-1999; 99US-0137502P.
XX PR 07-JUN-1999; 99US-0137724P.
XX PR 08-JUN-1999; 99US-0138094P.
XX PR 10-JUN-1999; 99US-0138540P.
XX PR 10-JUN-1999; 99US-0138847P.
XX PR 14-JUN-1999; 99US-0139119P.
XX PR 16-JUN-1999; 99US-0139452P.
XX PR 17-JUN-1999; 99US-0139453P.
XX PR 17-JUN-1999; 99US-0139492P.
XX PR 18-JUN-1999; 99US-0139454P.
XX PR 18-JUN-1999; 99US-0139455P.
XX PR 18-JUN-1999; 99US-0139456P.
XX PR 18-JUN-1999; 99US-0139457P.
XX PR 18-JUN-1999; 99US-0139458P.
XX PR 18-JUN-1999; 99US-0139459P.
XX PR 18-JUN-1999; 99US-0139460P.
XX PR 18-JUN-1999; 99US-0139461P.
XX PR 18-JUN-1999; 99US-0139462P.
XX PR 18-JUN-1999; 99US-0139463P.
XX PR 18-JUN-1999; 99US-0139750P.
XX PR 21-JUN-1999; 99US-0139763P.
XX PR 22-JUN-1999; 99US-0139817P.
XX PR 23-JUN-1999; 99US-0139899P.
XX PR 23-JUN-1999; 99US-0140353P.
XX PR 24-JUN-1999; 99US-0140354P.
XX PR 28-JUN-1999; 99US-0140695P.
XX PR 29-JUN-1999; 99US-0140823P.
XX PR 30-JUN-1999; 99US-0140991P.
XX PR 01-JUL-1999; 99US-0141287P.
XX PR 01-JUL-1999; 99US-0141842P.
XX PR 02-JUL-1999; 99US-0142154P.
XX PR 06-JUL-1999; 99US-0142390P.
XX PR 08-JUL-1999; 99US-0142803P.
XX PR 09-JUL-1999; 99US-0142920P.
XX PR 12-JUL-1999; 99US-0142977P.
XX PR 13-JUL-1999; 99US-0143542P.
XX PR 14-JUL-1999; 99US-0143624P.
XX PR 15-JUL-1999; 99US-0144005P.
XX PR 16-JUL-1999; 99US-0144085P.
XX PR 16-JUL-1999; 99US-0144086P.
XX PR 19-JUL-1999; 99US-0144325P.
XX PR 19-JUL-1999; 99US-0144331P.
XX PR 19-JUL-1999; 99US-0144332P.
XX PR 19-JUL-1999; 99US-0144333P.
XX PR 19-JUL-1999; 99US-0144334P.
XX PR 20-JUL-1999; 99US-0144335P.
XX PR 20-JUL-1999; 99US-0144352P.
XX PR 20-JUL-1999; 99US-0144632P.
XX PR 20-JUL-1999; 99US-0144884P.
XX PR 21-JUL-1999; 99US-0144814P.
XX PR 21-JUL-1999; 99US-0145086P.
XX PR 21-JUL-1999; 99US-0145088P.
XX PR 22-JUL-1999; 99US-0145085P.
XX PR 22-JUL-1999; 99US-0145087P.
XX PR 22-JUL-1999; 99US-0145089P.
XX PR 22-JUL-1999; 99US-0145132P.
XX PR 23-JUL-1999; 99US-0145145P.
XX PR 23-JUL-1999; 99US-0145218P.
XX PR 23-JUL-1999; 99US-0145224P.
XX PR 26-JUL-1999; 99US-0145276P.
XX PR 27-JUL-1999; 99US-0145913P.
XX PR 27-JUL-1999; 99US-0145918P.
XX PR 27-JUL-1999; 99US-0145919P.
XX PR 28-JUL-1999; 99US-0145951P.
XX PR 02-AUG-1999; 99US-0146386P.
XX PR 02-AUG-1999; 99US-0146388P.
```


PR	05-MAY-1999;	99US-0132485P.	PR	22-JUL-1999;	99US-0145087P.
PR	06-MAY-1999;	99US-0132486P.	PR	22-JUL-1999;	99US-0145089P.
PR	06-MAY-1999;	99US-0132487P.	PR	22-JUL-1999;	99US-0145192P.
PR	07-MAY-1999;	99US-0132863P.	PR	23-JUL-1999;	99US-0145145P.
PR	11-MAY-1999;	99US-0134256P.	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	99US-0134218P.	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	99US-0134219P.	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	99US-0134221P.	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	99US-0134370P.	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	99US-0134768P.	PR	27-JUL-1999;	99US-0145919P.
PR	19-MAY-1999;	99US-0134941P.	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	99US-0135124P.	PR	28-JUL-1999;	99US-0146386P.
PR	21-MAY-1999;	99US-0135353P.	PR	02-AUG-1999;	99US-0146388P.
PR	21-MAY-1999;	99US-0135652P.	PR	02-AUG-1999;	99US-0146389P.
PR	24-MAY-1999;	99US-0136021P.	PR	02-AUG-1999;	99US-0147038P.
PR	25-MAY-1999;	99US-0136022P.	PR	03-AUG-1999;	99US-0147204P.
PR	27-MAY-1999;	99US-0136392P.	PR	04-AUG-1999;	99US-0147302P.
PR	27-MAY-1999;	99US-0136782P.	PR	05-AUG-1999;	99US-0147132P.
PR	28-MAY-1999;	99US-0137222P.	PR	05-AUG-1999;	99US-0147260P.
PR	01-JUN-1999;	99US-0137522P.	PR	06-AUG-1999;	99US-0147303P.
PR	03-JUN-1999;	99US-0137502P.	PR	06-AUG-1999;	99US-0147416P.
PR	04-JUN-1999;	99US-0137724P.	PR	06-AUG-1999;	99US-0147493P.
PR	07-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147935P.
PR	08-JUN-1999;	99US-0138540P.	PR	10-AUG-1999;	99US-0148171P.
PR	10-JUN-1999;	99US-0138847P.	PR	11-AUG-1999;	99US-0148319P.
PR	14-JUN-1999;	99US-0139119P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148565P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148684P.
PR	16-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0149368P.
PR	17-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139455P.	PR	17-AUG-1999;	99US-0149436P.
PR	18-JUN-1999;	99US-0139456P.	PR	18-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149729P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139750P.	PR	26-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151066P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151080P.
PR	22-JUN-1999;	99US-0139899P.	PR	27-AUG-1999;	99US-0151080P.
PR	23-JUN-1999;	99US-0140353P.	PR	27-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140354P.	PR	31-AUG-1999;	99US-0151438P.
PR	24-JUN-1999;	99US-0140695P.	PR	30-AUG-1999;	99US-0151930P.
PR	28-JUN-1999;	99US-0140921P.	PR	01-SEP-1999;	99US-0151930P.
PR	29-JUN-1999;	99US-0142803P.	PR	07-SEP-1999;	99US-0152363P.
PR	30-JUN-1999;	99US-0141287P.	PR	10-SEP-1999;	99US-0153070P.
PR	01-JUL-1999;	99US-0141842P.	PR	13-SEP-1999;	99US-0153758P.
PR	01-JUL-1999;	99US-0142154P.	PR	15-SEP-1999;	99US-0154018P.
PR	02-JUL-1999;	99US-0142055P.	PR	16-SEP-1999;	99US-0154039P.
PR	06-JUL-1999;	99US-0142390P.			

```
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161320P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 65.5%; Score 1222; DB 3; Length 325;
Best Local Similarity 70.7%; Pred. No 2.9e-112;
Matches 224; Conservative 32; Mismatches 57; Indels 4; Gaps 3;

QY 39 IEPAILGVLCGSHAFHAFARFYKSYCSBEEIKKRFGIFVENLAFIRSTNRKDLST 98
Db 9 VEESVQILQSRHLSFARFTHRYGKYNVEEMKLRFSIPKENLDLIRSTNKKGLSYK 68

QY 99 LGINQFADLTWEFRNRLGAONCSATAGNHRFVDGVLPTROWREOGIVSPVKDQGS 158
Db 69 LGVNQFADLTWQBFQTKGAONCSATLKGSHKVTAAALPETKDWREDGIIVSPVKDQGG 128

QY 159 CGS-WTFSTTGALAAAYTQLTGS--TLSEQLVDCASAFNNFGC--GGLPSQAFYVYKNG 214
Db 129 CGSCWTFSTTGALAAAYHQAFAFGKGISLSEQLVDCAGAFNNYGCNGGLPSQAFYIKSNG 188

QY 215 GIDTEQTYPLGVNMGICNFKQENVGVKVIDSINITLGADELKHAAGLVLPVSVAFEVVK 274
Db 189 GLDTEKAYPTYGKDETCKPSAENVGVQLNSVNITLGADELKHAAGLVLPVSVIAFEVH 248

QY 275 GFNLKKGVSSTCGRDPMDNVHNAVLAGYGVGEDGIPYMLIKNSWGTWNGDNGYFMEL 334
Db 249 SFLYKSGVYVTDHSCGSTPMDNVHNAVLAGYGVGEDGVPYWLKNSWGDWGDGKGYFMEM 308

QY 335 GKNCMGVATCASYPIVA 351
Db 309 GKNCMGVATCASYPVVA 325
```

Search completed: March 29, 2004, 22:11:29
Job time : 89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:09:54 ; Search time 30 Seconds
(without alignments)
605.745 Million cell updates/sec

Title: US-10-087-714-2
Perfect score: 1865
Sequence: 1 MAKKLLFLLVLSVALS.....ELGKNMCGVATCASYPIVAV 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/aaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/aaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/aaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/aaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/aaa/PTUS_COMB.pep:*
6: /cgn2_6/prodata/2/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1300.5	69.7	378	4	US-09-325-932A-158
2	1252	67.1	361	3	US-09-120-365-75
3	1252	67.1	361	3	US-09-515-039-75
4	763	40.9	335	1	US-08-208-007A-10
5	763	40.9	335	1	US-08-330-121B-6
6	763	40.9	335	3	US-09-120-365-66
7	763	40.9	335	3	US-09-515-039-66
8	763	40.9	335	3	US-08-860-255A-6
9	763	40.9	335	4	US-08-915-095A-10
10	763	40.9	335	4	US-08-798-096-10
11	763	40.9	335	4	US-08-798-095A-10
12	763	40.9	335	4	US-09-953-956-10
13	763	40.9	335	4	US-08-553-125A-10
14	763	40.9	335	4	US-08-536-861-6
15	763	40.9	335	4	US-10-114-464-10
16	763	40.9	335	5	PCT-US95-13820-6
17	656.5	35.2	225	3	US-09-325-932A-162
18	656.5	35.2	225	3	US-09-120-365-67
19	656.5	35.2	225	3	US-09-515-039-67
20	603.5	32.4	339	3	US-09-120-365-62
21	603.5	32.4	339	3	US-09-120-365-80
22	603.5	32.4	339	3	US-09-515-039-62
23	603.5	32.4	339	3	US-08-515-039-80
24	594.5	31.9	323	3	US-09-120-365-61
25	594.5	31.9	323	3	US-09-515-039-61
26	594	31.8	500	4	US-09-325-932A-149
27	584.5	31.3	344	3	US-09-120-365-63

28	584.5	31.3	344	3	US-09-515-039-63	Sequence 63, Appl
29	584	31.3	386	4	US-09-325-932A-143	Sequence 143, Appl
30	583	31.3	374	3	US-08-821-994-67	Sequence 67, Appl
31	583	31.3	374	3	US-08-821-994-69	Sequence 69, Appl
32	581.5	31.2	467	3	US-08-821-994-83	Sequence 83, Appl
33	581	31.2	374	3	US-08-821-994-68	Sequence 68, Appl
34	580	31.1	370	4	US-09-325-932A-169	Sequence 169, Appl
35	573.5	30.8	443	3	US-08-821-994-65	Sequence 65, Appl
36	570	30.6	328	3	US-08-821-994-84	Sequence 84, Appl
37	562.5	30.2	351	1	US-09-500-651-2	Sequence 2, Appl
38	562.5	30.2	351	2	US-08-813-591-2	Sequence 2, Appl
39	562	30.1	374	3	US-08-821-994-70	Sequence 70, Appl
40	557	29.9	442	3	US-08-821-994-66	Sequence 66, Appl
41	556	29.8	380	3	US-09-120-365-76	Sequence 76, Appl
42	556	29.8	380	3	US-09-515-039-76	Sequence 76, Appl
43	556	29.8	380	3	US-08-860-255A-5	Sequence 5, Appl
44	554.5	29.7	457	3	US-09-120-365-72	Sequence 72, Appl
45	554.5	29.7	457	3	US-09-515-039-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-09-325-932A-158
; Sequence 158, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell death and their use in the modification of forestry plant level
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 158
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
; US-09-325-932A-158

Query Match 69.7%; Score 1300.5; DB 4; Length 378;
Best Local Similarity 70.4%; Pred. No. 8.7e-131;
Matches 247; Conservative 33; Mismatches 60; Indels 11; Gaps 5;

Qy	9	LLFLVLSALVALACFERDNPISVTQRPD----	SIEPAILGLVSGCRHAFHAFARFARYG	64
Db	31	LLLVAVVVSAAASFESNPILRF--PDGGLRDLESSIVQIVGTRHAFSFAFANRYG	87	
Qy	65	KSYGSEEEIKRGFGIFVENLAFIRSTNRKOLSYTLGINQADLTWEEFRTNRLGAAONCS	124	
Db	88	KRYETABEIKLRPFIFRENKLIRSTNNKGLPYTLGVNKFADWSWEEFRHRLGAAONCS	147	
Qy	125	ATAHGHRFVDGVLPTVRDWEQGI VSPVKDQSCGS-WTFSTTGALAAAYTQLTGS--T	181	
Db	148	ATTGKNHKLTDALPEMKDWEKGI VSPDKDQHGCGSWTFSTTGALAAAYHQAQKQIS	207	
Qy	182	LSEQQLVDCASAFNNFGC--GGLPSQAFYKYKNGIDITEQTYPYLVGMGICNFQENGV	240	
Db	208	LSEQQLVDCAGAFNNPGCSGGLPSQAFYKYKNGIDITEEAYPYTAVDGSKCFSDNVGV	267	
Qy	241	KVIDSINITGADELKHAIVLPVSVAVKVGNNLYKKGVYSSDTCCGRDPMVNHAV	300	
Db	268	QVLDSVINITGADELKHAIVLPVSVAVKVGNNLYKKGVYSSDTCCGRDPMVNHAV	327	
Qy	301	LAVGVGVEDGIPYWLIIKNSWGTWNGDNGYFPMELGKNMCGVATCASYPIVA	351	
Db	328	LAVGVGVEDGVFPLIIKNSWGTWNGDNGYFPMELGKNMCGVATCASYPIVA	378	

RESULT 2

US-09-120-365-75
; Sequence 75, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-14479
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 75
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Aleurain
US-09-120-365-75

Query Match 67.1%; Score 1252; DB 3; Length 361;
Best Local Similarity 67.3%; Pred. No. 1.3e-125;
Matches 241; Conservative 42; Mismatches 65; Indels 10; Gaps 7;

QY 3 AKLLFLLFLVSALSVALA---GPEEDNPISVTPQDPS--IEPAILGVLGSCRHAFHFA 58
DB 4 ARVLLALAVLATAAVALVASSSSFADSNPIRPVTDRAASTLESVALGALGRTHALRFA 63
QY 59 FARYYKSGYSGEIEIKKRFIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEPRNRLG 118
DB 64 FAVRYKSGYESAAEVRFRIFSESLEBVRSTNRKGLPYRLGINRFSMDSWEEFQATRLG 123
QY 119 AAQCSATAHGNHRFVD--GVLPVTRDWEQIGIVSPVKDQSGCS--WTFSTTGALAAAYTQ 176
DB 124 AAQCSATLGNHLMRDAALPETKDWREDDGIVSPVKNAHCSCWTFSTTGALAAAYTQ 183
QY 177 LTGS--TLSEQQLVDCSAFNNFGC--GGLPSQAFYVKYNGIGIDTEQTYPLGVNMGICNF 233
DB 184 ATGKNISLSEQQLVDCAGFNNFGCGGLPSQAFY--QYNGGIDTESPYKGVNGVCHY 242
QY 234 KOENVGVKVIDSINTILGADELKHAHLVPRVSVAFVVKGNLYKKGVSSTDCGRDP 293
DB 243 KAENAAVQVLDVSNITLNAEDELKNAVLVPRVSVAFQVIDGFRQYKSGVITSDHCSTTP 302
QY 294 MDVNHAVLAGVGVEDGIPYMLIKNSWGTNMGDNGYFKMELGKNKMGVATCASYPIVA 351
DB 303 DDVNHAVLAGVGVENGVPYMLTKNSWADGNDNGYFKMENGKNKMCATCATCASYPVVA 360

RESULT 3
US-09-515-039-75
; Sequence 75, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-14479
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 75
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Aleurain
US-09-515-039-75

Query Match 67.1%; Score 1252; DB 3; Length 361;
Best Local Similarity 67.3%; Pred. No. 1.3e-125;
Matches 241; Conservative 42; Mismatches 65; Indels 10; Gaps 7;

QY 3 AKLLFLLFLVSALSVALA---GPEEDNPISVTPQDPS--IEPAILGVLGSCRHAFHFA 58

DB 4 ARVLLALAVLATAAVALVASSSSFADSNPIRPVTDRAASTLESVALGALGRTHALRFA 63
QY 59 FARYYKSGYSGEIEIKKRFIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEPRNRLG 118
DB 64 FAVRYKSGYESAAEVRFRIFSESLEBVRSTNRKGLPYRLGINRFSMDSWEEFQATRLG 123
QY 119 AAQCSATAHGNHRFVD--GVLPVTRDWEQIGIVSPVKDQSGCS--WTFSTTGALAAAYTQ 176
DB 124 AAQCSATLGNHLMRDAALPETKDWREDDGIVSPVKNAHCSCWTFSTTGALAAAYTQ 183
QY 177 LTGS--TLSEQQLVDCSAFNNFGC--GGLPSQAFYVKYNGIGIDTEQTYPLGVNMGICNF 233
DB 184 ATGKNISLSEQQLVDCAGFNNFGCGGLPSQAFY--QYNGGIDTESPYKGVNGVCHY 242
QY 234 KOENVGVKVIDSINTILGADELKHAHLVPRVSVAFVVKGNLYKKGVSSTDCGRDP 293
DB 243 KAENAAVQVLDVSNITLNAEDELKNAVLVPRVSVAFQVIDGFRQYKSGVITSDHCSTTP 302
QY 294 MDVNHAVLAGVGVEDGIPYMLIKNSWGTNMGDNGYFKMELGKNKMGVATCASYPIVA 351
DB 303 DDVNHAVLAGVGVENGVPYMLTKNSWADGNDNGYFKMENGKNKMCATCATCASYPVVA 360

RESULT 4
US-08-208-007A-10
; Sequence 10, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSES: CARRELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSES: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A
; FILING DATE: March 8, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: No. 5501969e
; FILING DATE: No. 5501969e
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-208-007A-10

Query Match 40.9%; Score 763; DB 1; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;
QY 54 FHFARFARYKSGYSGEIEIKKRFIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEPR 113

Wed Mar 31 08:14:35 2004

```
US-09-515-039-66
; Sequence 66, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapien cathepsin H
US-09-515-039-66

Query Match 40.9%; Score 763; DB 3; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY 54 FHFAFPARRYKSGYSGSEEEIKKRGIFVENLAFIRSTNRKDLSTYLTGINQFADLTWEEFR 113
Db 33 PHFKSWMSKHKRTY-STEEYHRLQTFASNWRKINAHNNGNHTFKMALNQFSDMSFAEIK 91
QY 114 TNRL-GAQNCSATAHGNHFRVDGLPVTRDREOG-IVSPVKDQSCGS-WTFSTTGAL 170
Db 92 HKYLWSEFQNCSAT-KSNYLRTGTPYPSVDWRKKNFVSPVKNGACGSCWTFSTTGAL 150
QY 171 EAAVTQLTGSTLS--EQLVDCASAFNNFGC-GGLPSQAFYVKYKNGIDTEQTPYLG 227
Db 151 ESAIAIATGKMLSLAEQQLVDCADFNNYCGQGLPSQAFYILYKNGIMGEDTYPYQ 210
QY 228 MGICNFQKQENVGVKVIDSINITLGADELKHAUGLVRPVSVAPEVVKGFNLYKGYSSD 287
Db 211 DGYCKFQPKAIGFVKDVAIITYDEEAMVEAVALYNPVSFAFEVTDQFMWRTGIYSST 270
QY 288 TCGRDPMDVNHAVLAGVGVEDGIPYWLKNSWGTNGDNGYFQWELGKNKCGVATCASY 347
Db 271 SCHKTPDKVNHAVLAGVGEKNGIPYLVKNSWGPQGMNGYFLIERGKNMCGLAACASY 330
QY 348 PIVAV 352
Db 331 PIPLV 335

RESULT 8
US-08-860-255A-6
; Sequence 6, Application US/08860255A
; Patent No. 6274336
; GENERAL INFORMATION:
; APPLICANT: Abdel-Mequid, Sherin
; APPLICANT: Desjarlais, Renee
; APPLICANT: Janson, Cheryl
; APPLICANT: Smith, Ward
; APPLICANT: Zhao, Baoquang
; TITLE OF INVENTION: Method of Inhibiting Cathepsin K
; FILE REFERENCE: P50574-X1
; CURRENT APPLICATION NUMBER: US/08/860,255A
; CURRENT FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: 60/008,108
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: 60/007,473
; PRIOR FILING DATE: 1995-11-22
; PRIOR APPLICATION NUMBER: 60/008,992
; PRIOR FILING DATE: 1995-12-21
; PRIOR APPLICATION NUMBER: 60/013,748
; PRIOR FILING DATE: 1996-03-20
; PRIOR APPLICATION NUMBER: 60/013,764
; PRIOR FILING DATE: 1996-03-20
; PRIOR APPLICATION NUMBER: 60/013,747

US-08-915-095A-10
; Sequence 10, Application US/08915095A
; Patent No. 6383793
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PE10704
; CURRENT APPLICATION NUMBER: US/08/915,095A
; CURRENT FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-915-095A-10

Query Match 40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY 54 FHFAFPARRYKSGYSGSEEEIKKRGIFVENLAFIRSTNRKDLSTYLTGINQFADLTWEEFR 113
```

33	PHFKSMXKHKXTY-STBEYHRLQTFASNNWRKINAHNNNGNHTFKALANQFSDMSFAETK	91
114	TNRL-GAAGNCSATAHGNHRFVDGVLPVTRDWRQEQ-IVSPVKDQSGSGS-WTFSTTGAL	170
92	HXYLWSEPNQCSAT-KSNYLRTGTPYPSPVDWRKKGNFVSPVKNQAGCSGCTFTTGAL	150
171	EAAYTQLTGTSLTS--EQLVPCASAFNNFGC-GGLPSQAFYKVKYNGIDTBTQYPLVG	227
151	ESGAIAIATGKMLSLAEQQLVDCADFNNGYCGQLPSQAFEYILYKNGIMGEDTYPYQKG	210
228	MGICNFKQENVGKVIIDSTNITLGAEDDLKHAGLVAPVSAVEVVKFENLYKKGVSSD	287
211	DGCFKPQPKAKTGFVKDVAANIITIDEEANWEAVALYNPVSFAFEVTDQFMFKRTGIYSST	270
288	TGCRDPMDVNHAFLAVGCVGEDGIPYMLIKNSWGTNGNDNGYFKMELGKMKCVATCASY	347
271	SCHKTPDKNHAVLAVGGEKNGIPYIVKNSWGPQMGNGYFLIERGKMKCGLAACASY	330
348	PIVAV	352
331	PIPLV	335

```

RESULT 10
US-08-798-096-10
; Sequence 10, Application US/08798096
; Patent No. 6387682
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2
; CURRENT APPLICATION NUMBER: US/08/798,096
; CURRENT FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-096-10

```

Query Match 40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred.No.3.le-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7

Qy	54	FHFAFARFYKXSYGSEBEIKKRGFOGVFVENLAFTRSNVRKDLSVYTGINOQFADLTWEEPR	113
Db	33	FHFKSWMSKHRTY-STBEYHRLQTFASNWRKINAHNNGHNTFOKALNQFSDMSFAEIK	91
Qy	114	TNRL-GAAQNCSATAHGNHRFVDGVLVPTRDWREQ-IVSPVVKDOQSCGS-WTFSTTGAL	170
Db	92	HKYLHSEPNQCSAT-KSNYLRTGPPYPSVDWRKKGNFVSPVKNOGACGSCWTFSTTGAL	150
Qy	171	EAAVTLTGSTLS--EQQLVDCASFNFGC-GGLPSQAFYKYKNGGIDTEQTYPLGV	227
Db	151	ESAIAIATCKMLSLAEQOLVPCADPNFYCGGGLPSQAFEYLYNKGINGEDTYPYQK	210
Qy	228	MGICNFPQENVGVRKVIDSINITTLGAEDBELHAGVLVFPVSAVEVVKGFNLYKKGYSSD	287
Db	211	DGYCKFPQCAIGFVKVDVANITIVDEEAMVEAVALYNPVSFAPEVTQDFMYRTGIYSST	270
Qy	288	TGCRDPMYDHNHVAIVAGYVEDGTPYMLIKNSWCTNNGDNGYFKMELGKNMCCGVATCASY	347
Db	271	SCHKTDPKHNHVAIVAGYGEKNGIPYMLIKNSWGPQWGMNGYFLIERGKNMCCGLAACASY	330
Qy	348	PIVAV	352
Db	331	PIPLV	335

RESULT 11
US-08-798-095A-10

```

; Sequence 10, Application US/08798095A
; Patent No. 6423507
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D3
; CURRENT APPLICATION NUMBER: US/08/798, 095A
; CURRENT FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-798-095A-10

```

[illegible]

```

RESULT 12
US-09-953-956-10
; Sequence 10, Application US/09953956
; Patent No. 6475487
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-10

```

```
Query Match      40.9%; Score 763; DB 4; Length 335;  
Best Local Similarity 52.1%; Pred.No. 3.le-73;  
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;
```

114 TNRL-GAAQNCATAGHNRHFRVDPVLPVTRDWEQ-IVSPVKDQSGS-WTFSTTGAL 170
92 HKYLWSEPQNCAT-KSNYLRTGTPPPSVDRKKNFVSPVKQACGSCWTFSTTGAL 150
171 EAAATLTGSLT--EQQLVDCASAFNFGC-GGLPSQAFYVYKNGGIDTQTPYLV 227
151 ESAIATATGKMLSLAEQQLVDCQDFNNYCGGGLPSQAFYLYNKGIMGSDTYPYQK 210
228 MGICNFKQENVGVKVIDSINITLGADELKHAVALVPRVSVAFVYVYKNGGIDTYPYQK 287
211 DGYCKFQPKAIGFVKDVAITTYDEAMVEALYNFVSFAFVTDQFMVYRTGIYST 270
288 TGRDPMVNHAVLAVGYVEDGIPYWLKNSGNTWNGDGYFKMELGKMGCVATCASY 347
271 SCHKTPDKVNHAVLAVGYGKNGIPYVKNWGPQWNGYFLIERGKMGCLAACASY 330
348 PIVAV 352
331 PIPLV 335

RESULT 13
US-08-553-125A-10
; Sequence 10, Application US/0853125A
; Patent No. 6475766
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: P107D1
; CURRENT APPLICATION NUMBER: US/08/553.125A
; CURRENT FILING DATE: 1995-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-553-125A-10

Query Match 40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;
QY 54 PHEAFARYGKSYGSEEEIKRFGIFVENLAFIRSTNRKDLSTVLGINQFADLTWEEFR 113
DB 33 FHFKSWMSKHKRTY-STEEYHRLQTFASNRKINAHNNGNHTFMAINQFSDMSFAEIK 91
QY 114 TNRL-GAAQNCATAGHNRHFRVDPVLPVTRDWEQ-IVSPVKDQSGS-WTFSTTGAL 170
DB 92 HKYLWSEPQNCAT-KSNYLRTGTPPPSVDRKKNFVSPVKQACGSCWTFSTTGAL 150
QY 171 EAAATLTGSLT--EQQLVDCASAFNFGC-GGLPSQAFYVYKNGGIDTQTPYLV 227
DB 151 ESAIATATGKMLSLAEQQLVDCQDFNNYCGGGLPSQAFYLYNKGIMGSDTYPYQK 210
QY 228 MGICNFKQENVGVKVIDSINITLGADELKHAVALVPRVSVAFVYVYKNGGIDTYPYQK 287
DB 211 DGYCKFQPKAIGFVKDVAITTYDEAMVEALYNFVSFAFVTDQFMVYRTGIYST 270
QY 288 TGRDPMVNHAVLAVGYVEDGIPYWLKNSGNTWNGDGYFKMELGKMGCVATCASY 347
DB 271 SCHKTPDKVNHAVLAVGYGKNGIPYVKNWGPQWNGYFLIERGKMGCLAACASY 330
QY 348 PIVAV 352
DB 331 PIPLV 335

RESULT 14
US-08-536-861-6
; Sequence 6, Application US/08536861
; Patent No. 6544767
; GENERAL INFORMATION:

APPLICANT: Bromme, Dieter
APPLICANT: Okamoto, Kathleen
TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/536.861
FILING DATE: 02-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-60261-1/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-536-861-6

Query Match 40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;
QY 54 PHEAFARYGKSYGSEEEIKRFGIFVENLAFIRSTNRKDLSTVLGINQFADLTWEEFR 113
DB 33 FHFKSWMSKHKRTY-STEEYHRLQTFASNRKINAHNNGNHTFMAINQFSDMSFAEIK 91
QY 114 TNRL-GAAQNCATAGHNRHFRVDPVLPVTRDWEQ-IVSPVKDQSGS-WTFSTTGAL 170
DB 92 HKYLWSEPQNCAT-KSNYLRTGTPPPSVDRKKNFVSPVKQACGSCWTFSTTGAL 150
QY 171 EAAATLTGSLT--EQQLVDCASAFNFGC-GGLPSQAFYVYKNGGIDTQTPYLV 227
DB 151 ESAIATATGKMLSLAEQQLVDCQDFNNYCGGGLPSQAFYLYNKGIMGSDTYPYQK 210
QY 228 MGICNFKQENVGVKVIDSINITLGADELKHAVALVPRVSVAFVYVYKNGGIDTYPYQK 287
DB 211 DGYCKFQPKAIGFVKDVAITTYDEAMVEALYNFVSFAFVTDQFMVYRTGIYST 270
QY 288 TGRDPMVNHAVLAVGYVEDGIPYWLKNSGNTWNGDGYFKMELGKMGCVATCASY 347
DB 271 SCHKTPDKVNHAVLAVGYGKNGIPYVKNWGPQWNGYFLIERGKMGCLAACASY 330
QY 348 PIVAV 352
DB 331 PIPLV 335

RESULT 15
US-10-114-464-10
; Sequence 10, Application US/10114464
; Patent No. 6680375
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN

```
FILE REFERENCE: PF107DS
CURRENT APPLICATION NUMBER: US/10/114,464
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 08/553,125
PRIOR FILING DATE: 1995-11-07
PRIOR APPLICATION NUMBER: 08/208,007
PRIOR FILING DATE: 1994-03-08
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 335
TYPE: PRT
ORGANISM: Homo sapiens
US-10-114-464-10

Query Match      40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY 54 FHFAFARYGKSYGSEEEIKKGFIVENLAFIRSTNRKDLSTYTLGINOFADLTWEPR 113
DB 33 FHFKSWMKHKRTY-STEEYHRLQTFASNRKKNHAHNNHNTFKNALNQFSDMSFAIK 91

QY 114 TNRL-GAAQNCSATAGNHRFVDGLPVTRDREQ-IVSPVKDQSCGS-WTFSTTGAL 170
DB 92 HKYLNSEPCNSAT-KSNYLRGTGYPPEVDNRKKNFVSPVKQAGCGSCWTFSTTGAL 150

QY 171 EAAVTQLTGSTLS--EQQLVDCASAPNFGC--GGLPSQAFHYVKNKGIDTQTPYPLGV 227
DB 151 ESAIAIATGKMLSLAEQQQLVDCADQDFNNYCGQGLPSQAFYILYNKIMGIDTYPYQK 210

QY 228 MGICNFKQENVGVKVIDSINITLGADELKHAVGLVRPVSAFVGVVKNLYKGVSSD 287
DB 211 DGYCKEQPKAIGFVKDVANIITYDEAMVEAVALYNPVSAFVTDQFMRYRGIYST 270

QY 288 TCRDPMVNHAVLAVGYVEDGIPYWLKNSWGTNWDNGYFKMELGNMCGVATCASY 347
DB 271 SCHKTPDKYNHAVLAVGYGKNGIPYWIIVKNSWGPQWGMNGYFLIERGNMCGLAACASY 330

QY 348 PIVAV 352
DB 331 PIPLV 335
```

Search completed: March 29, 2004, 22:16:57
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:15:19 ; Search time 332 Seconds
(without alignments)
277.425 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAAKLFFLLFLVSALVAL.....ELGKNMGVATCASYPIVAV 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1865	100.0	352	14	US-10-087-714-2
2	1323	70.9	363	14	Sequence 2, Appli
3	1321.5	70.9	368	12	Sequence 24, Appl
4	1320.5	70.8	368	12	Sequence 56425, A
5	1319.5	70.8	368	12	Sequence 59088, A
6	1319.5	70.8	368	12	Sequence 42679, A
7	1310.5	70.3	340	12	Sequence 69552, A
8	1302.5	69.8	360	14	Sequence 59728, A
9	1300.5	69.7	378	14	Sequence 23, Appl
10	1286	69.0	360	14	Sequence 158, App
11	1280	68.6	353	12	Sequence 22, Appl
12	1280	68.6	353	12	Sequence 195993, A
13	1272.5	68.2	360	14	Sequence 4426, A
14	1261.5	67.6	362	14	Sequence 20, Appl
15	1255.5	67.3	362	14	Sequence 25, Appl
					Sequence 21, Appl

16	1252	67.1	308	12	US-10-425-114-51874
17	1249	67.0	361	12	Sequence 51874, A
18	1236.5	66.3	357	12	Sequence 49461, A
19	1148	61.6	355	14	Sequence 283471, A
20	852.5	45.7	193	12	Sequence 283, App
21	845.5	45.3	277	14	Sequence 46437, A
22	838	44.9	201	12	Sequence 280, App
23	828.5	44.4	185	12	Sequence 42955, A
24	780.5	41.8	168	12	Sequence 57524, A
25	763	40.9	335	9	Sequence 46575, A
26	763	40.9	335	13	Sequence 10, Appl
27	763	40.9	335	14	Sequence 6, Appl
28	759	40.7	364	9	Sequence 478, Appl
29	656.5	35.2	225	14	Sequence 162, App
30	609	32.7	470	14	Sequence 278, App
31	598.5	32.1	381	12	Sequence 43777, A
32	594.5	31.9	452	14	Sequence 276, App
33	594	31.8	500	14	Sequence 149, App
34	592.5	31.8	497	12	Sequence 65121, A
35	590.5	31.7	494	12	Sequence 49414, A
36	590.5	31.7	495	12	Sequence 67151, A
37	590.5	31.7	495	12	Sequence 67152, A
38	589	31.6	464	15	Sequence 7, Appli
39	589	31.6	479	12	Sequence 4438, A
40	589	31.6	479	12	Sequence 55418, A
41	584	31.3	386	14	Sequence 143, App
42	584	31.3	478	12	Sequence 44927, A
43	584	31.3	478	12	Sequence 51948, A
44	584	31.3	478	12	Sequence 53490, A
45	583.5	31.3	480	12	Sequence 47349, A

ALIGNMENTS

RESULT 1

US-10-087-714-2
; Sequence 2, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; APPLICANT: Dixon, Richard A.
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMC10099
; CURRENT APPLICATION NUMBER: US/10/087,714
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 09/462,576
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US98/14895
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/052,604
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: 60/272,415
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Vanilla planifolia
US-10-087-714-2

Query Match	100.0%	Score 1865	DB 14	Length 352
Best Local Similarity	100.0%	Pred. No. 3.5e+188		
Matches 352	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	MAAKLFFLLFLVSALVALAGFEEDNP	IRSVTORPDSIEPAILGVLSGRHAFHARFA	60
Db	1	MAAKLFFLLFLVSALVALAGFEEDNP	IRSVTORPDSIEPAILGVLSGRHAFHARFA	60
Qy	61	RRYKSGVSEETKKEFGIVFNLAIRSNR	KDLSVTGLGINQFADLTWEEPTNRLGAA	120

Db 61 RRYKSYGSEETIKKREGIFVENLAFIRSTNRKDLSTYTLGINOAFDLTWEEFRTNRLGAA 120
QY 121 QNCSATAGHNRHFRVDGVLPTVRDREGIVSPVKDQSGSWTFTTGALEAAATOLTGS 180
Db 121 QNCSATAGHNRHFRVDGVLPTVRDREGIVSPVKDQSGSWTFTTGALEAAATOLTGS 180
QY 181 TLSEQQQLVDCASAFNFGCGGLPSQAFYKYNKGIDTQTYPLGVGMGICNFKQENGV 240
Db 181 TLSEQQQLVDCASAFNFGCGGLPSQAFYKYNKGIDTQTYPLGVGMGICNFKQENGV 240
QY 241 KVIDSINITGADELKHAHGLVRPVSVAFEVVKGNLYKKGYSDDTCGRDMDVNHAV 300
Db 241 KVIDSINITGADELKHAHGLVRPVSVAFEVVKGNLYKKGYSDDTCGRDMDVNHAV 300
QY 301 LAVGYGVEDGIPYWLKNSGWTNGDNGYFKMELGKMGCVATCASYPIVAV 352
Db 301 LAVGYGVEDGIPYWLKNSGWTNGDNGYFKMELGKMGCVATCASYPIVAV 352

RESULT 2

US-10-087-714-24
; Sequence 24, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; APPLICANT: Dixon, Richard A.
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; CURRENT APPLICATION NUMBER: US/10/087,714
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 09/462,576
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US98/14895
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/052,604
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: 60/272,415
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 24
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Zea mays
US-10-087-714-24

Query Match 70.9%; Score 1323; DB 14; Length 363;
Best Local Similarity 70.8%; Pred. No. 7.2e-131;
Matches 257; Conservative 30; Mismatches 64; Indels 12; Gaps 6;

QY 1 MAKKLLPFLFL----VSALSVALAGFEEDNPISVTQRPDS-IEPAILGVLSGCRHAF 54
Db 1 MAHRIILLVAVAATSAVAASAGDDSDNPPIRPTVTDRAASALESTVFAALGRTDAL 60
QY 55 HFAPFARYGKSYGSEEEIKKRGIFVENLAFIRSTNRKDLSTYTLGINOAFDLTWEEFRT 114
Db 61 RFAPFARYGKSYGSEAAEVHKKRFRIFSSQLVRSNRKGLSYRLGYNRFADMSWEEFRA 120
QY 115 NRLGAQNCSTAHGNHRF--VDGVLPTVRDREGIVSPVKDQSGCS-WTFSTTGALE 171
Db 121 TRLGAQNCSTALTGNHRMAAAVALPETKDWREDGIVSPVKDQSGCSWTFSTTGALE 180
QY 172 AAYTOLTGS--TLSEQQQLVDCASAFNFGC--GGLPSQAFYKYNKGIDTQTYPLGVGM 228
Db 181 AAYTQATKPISSLSEQQQLVDCGFAPNFGCGGLPSQAFYKYNKGIDTQTYPLGVGM 240
QY 229 GICNFKQENGVKVIDSINITGADELKHAHGLVRPVSVAFEVVKGNLYKKGYSDDT 288
Db 241 GICNFKQENGVKVIDSINITGADELKHAHGLVRPVSVAFEVVKGNLYKKGYSDDT 300
QY 289 CGRDPMDVNHAVLAVGYGVEDGIPYWLKNSGWTNGDNGYFKMELGKMGCVATCASYP 348

Db 301 CGTTPMDVNHAVLAVGYGVEDGIPYWLKNSGWTNGDNGYFKMELGKMGCVATCASYP 360
QY 349 IVA 351
Db 361 IVA 363
RESULT 3
US-10-425-114-56425
; Sequence 56425, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)E
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56425
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB189-021-F2_FLI.pep
US-10-425-114-56425

Query Match 70.9%; Score 1321.5; DB 12; Length 368;
Best Local Similarity 71.7%; Pred. No. 1.1e-130;
Matches 258; Conservative 27; Mismatches 66; Indels 9; Gaps 6;

QY 1 MAKKLLPFLFLVSALSVALA--GFEEDNPISVTQRPDS-IEPAILGVLSGCRHAFHA 57
Db 9 MAPRELLVAVVLAATAAANSAGPADSNPIRPTVTDRAASALESTVFAALGRTDALRFA 68
QY 58 REAPRYGKSYGSEEEIKKRGIFVENLAFIRSTNRKDLSTYTLGINOAFDLTWEEFRTNRL 117
Db 69 RFAPRYGKSYGSEAAEVHKKRFRIFSSQLVRSNRKGLSYRLGYNRFADMSWEEFRTAL 128
QY 118 GAAQNCSTAHGNHRF--VDGVLPTVRDREGIVSPVKDQSGCS-WTFSTTGALEAAY 174
Db 129 GAAQNCSTALTGNHRMAAAVALPETKDWREDGIVSPVKDQSGCSWTFSTTGALEAAY 188
QY 175 TOLTGS--TLSEQQQLVDCASAFNFGC--GGLPSQAFYKYNKGIDTQTYPLGVGMGIC 231
Db 189 TQATKPISSLSEQQQLVDCGFAPNFGCGGLPSQAFYKYNKGIDTQTYPLGVGMGIC 248
QY 232 NFKQENGVKVIDSINITGADELKHAHGLVRPVSVAFEVVKGNLYKKGYSDDTCGR 291
Db 249 KFKENGVKVIDSINITGADELKHAHGLVRPVSVAFEVVKGNLYKKGYSDDTCGR 308
QY 292 DMDVNHAVLAVGYGVEDGIPYWLKNSGWTNGDNGYFKMELGKMGCVATCASYPIVA 351
Db 309 TMDVNHAVLAVGYGVEDGIPYWLKNSGWTNGDNGYFKMELGKMGCVATCASYPIVA 368

RESULT 4

US-10-425-114-59088
; Sequence 59088, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59088
; LENGTH: 368
; TYPE: PRI
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700048693_FJI.psp
US-10-425-114-59088

```

Query Match	70.8%	Score 1320.5	DB 12	Length 368
Best Local Similarity	71.7%	Pred. No. 1.4e-130		
Matches 258	Conservative 26	Mismatches 67	Indels 9	Gaps 6

QY	1	MAAKLLPFLFLVLSALVALA--GFEDNPNRSVTQRPDS--IEPAILLVLGSCRHAFHEA	57
DB	9	MAPRLLLVLVALLAATAAANAAGFADSNPIRPVTDRAASALESTVFAALGRTRDAURFA	68
QY	58	RFARRYKSYGSBEEIKRQIGFVFNELAFIRSTNRKDLSTVTLGINQFADLTWEEFRNRL	117
DB	69	RFARVYKSYESAEEVHKRFRIFSESILQVRSTNRKGLSYRLGINRRADMSWEEFRATRL	128
QY	118	GAQNCSATAHGNHRF--VDGVLPVTRDWRQGIIVSPVKDQSGCS--WTFTTGALEAAV	174
DB	129	GAQNCSATITGNHRMFAAAVALPETKDRWDGIVSPVTQNGHCGSCWTFTTGALEAAV	188
QY	175	TQLTGS--TLSEOLVDCAAFNFGC--GGLPQAEEVVKYNGGIDTQRTQYVYLGVWGIC	231
DB	189	TQATGRPISUSEOLVDCGFANFGNCGGLPQAFFIYKNGGLDTEESYPQGVNGIC	248
QY	232	NFXQENVGVKVIDSINITLTGAEDBLKHAVLVRPVSVAFEVVKGFNLYKXGVSSDTCGR	291
DB	249	KFXKNVGVKVLDSVNITLTGAEDBLKDAVLVRPVSVAFEVITGFRLYKSGVYTSDHCGT	308
QY	292	DPMDVNHAVLAVGVGEDGTPYMLIKNSGTNNGDNGYFKRELGNKNCGVATCASYPIVA	351
DB	309	TPMDVNHAVLAVGVGEDGTPYMLIKNSWADWDGDEGYFKNEMGNKNCGVATCASYPIVA	368

```

RESULT 5
US-10-425-114-42679
; Sequence 42679, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Jinhua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313) B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 42679
; LENGTH: 368
; TYPE: PRI
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700443856_FLI.pep
US-10-425-114-42679

```

Query Match 70.8%; Score 1319.5; DB 12; Length 368;
Best Local Similarity 71.4%; Pred. No. 1.7e-130;
Matches 257; Conservative 27; Mismatches 67; Indels 9; Gaps 6;

QY 1 MAAKLLFFLLFLVSALSVALA--GFEDNPIRSVTQRPSD-IETALLGVIGSCRHAFHA 57

```

Db      9  MAPRLVLAVVALAAATAAANAANGFADSNPIRPTVTDRAASALESTVFAALGKTRDALRPA 68
Qy      58  RFARRYKSGXSEBEEIKKRFQIGVFNELAFIRSTNRKOLSYTLGINQFADLTWEEFTRNL 117
Db      69  RFAVRYKGSYESAAAEVHKRFPRIFESLQQLVRSTNRKGLSYRLGINRFAFADMSWEEFRATRL 128
Qy     118  GAAQNCSATAGNHRF--VDGVLPTVRDWRQGIQVSPVKDOGSCGS-WTFSTTGALEAA 174
Db     129  GAAQNCSATUTGNHRMRAAAVALPETKDWREDGIVSPVKNGHQHCSGWTFTTGALEAA 188
Qy     175  TQLTGS--TLSEQQLVPCASAFNNFGC-GGLPSQAFEXYKVKNGIGDITQTPYPLGVMGIC 231
Db     189  TQATGKFISEQLIDCGFAFNFGCGGLPSQAFEVIKYNGLDTEESYPQGVNGIC 248
Qy     232  NFKQENVGVKVIDSINTILGAEDLKHAVGLVRPVSAFAFVVKGNLYKKGYVSSDTCGR 291
Db     249  KFNENYGVKVLDSVNTILGAEDLKDQAVGLRFPVSAFEVITGRLYKSGVYTSDHCGT 308
Qy     292  DPMDVNAVLAVGVGEDGIPYLLIKNSWGTNMGDNGYFKMKELKQMGVATCASYPIVA 351
Db     309  TPMDVNAVLAVGVGEDGIPYLLIKNSWGDWDGEGYFKMKEMKQMGVATCASYPIVA 368

RESULT 6
US-10-425-114-69552
; Sequence 69552, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingtong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69552
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73058A12_F11.pep
US-10-425-114-69552

```

Query Match 70.8%; Score 1319.5; DB 12; Length 368;
Best Local Similarity 71.4%; Pred. No. 1.7e-130;
Matches 257; Conservative 27; Mismatches 67; Indels 9; Gaps 6;

Qy	1	MAAKULLFFLLFVLVSALSVALA--GFEBDNPIRSTVQRPDS-IEPAILGVLGSCRHAFHFA	57
Db	9	MAPRELLLVAVALAATAAANSFGADSNPIRPVPTDRAASALESTVPAALGRTRDALRFA	68
Qy	58	RFARVYKGSYGSEEBIKKRFGIFVENLAFIRSTNRKDSLTYTLGINOFADLTWEEFRTNEL	117
Db	69	RFARVYKGSYSAAEVHVKRFIFSESILQVLRSTNRKGLSYRLGINFADMSWEEFPAATL	128
Qy	118	GAQNCSTATAGNHRFP--VDGVLPTVTDWRBQIGLVSVPVKQSGGS-WFSTTGALAEAY	174
Db	129	GAQNCSTATLGNHRMRAAAVALPETKDWREDGIVSPVKQSGCSWTFTSTTGALAEAY	188
Qy	175	TQLTGS--TLSEBQQLVDCASAFNFGC--GGLPSQAFVYVKNYGGIDTEQTPYLVGMVGC	231
Db	189	TQATGKPLSEBQQLIDCGFAFNFGCGGLPSQAFEVIKYNGGLDTEESYPYQGVNGIC	248
Qy	232	NFKQENVGKVLDSINITLGADELKHAVGLRVPVSAFVWGVNGLYKKGVTSSDTCGR	291
Db	249	KPKENVGKVLDSVNIITLGADELKDAVGLRVPVSAFVITGFLYKSGVTVSDHCGT	308
Qy	292	DPMDVNVHVLAVGYGVEDGIEPYWLIKNSWGTNMGDNGYFKEVLGKQMCVATCASYPIVA	351

Db 88 KYETAEELKLPREFENLKIIRSTNKGGLPYTLGVNKFADSWEBFRHRLGHAQNC 147
Qy 125 ATAHGNHREVDGVLVPTDRDREQGVSPVKDQSCGS-WTFSTTGALAAAYTQLTGS--T 181
Db 148 ATTGNHKLTDALPEMDKREKIVSPKIDQCHGCGSCWTFTSTTGALAAAYHQAQKQIS 207
Qy 182 LSEQQLVDCASAFNNFGC--GGLPSQAFVYKNGGIDTEQTPYLVGMGICNFKQENYGV 240
Db 208 LSEQQLVDCAGAFNFGCGGLPSQAFVYKNGGIDTEEAYPTAVDQCKFSADNVGV 267
Qy 241 KVIDSINITLGADELKHAVALVRPVSVAFEVVKGFNLYKKGVSDDTCGRDPMVNHAV 300
Db 268 QVLDSVNITLGADELKHAVALVRPVSVAFEVVKGFNLYKKGVSDDTCGRDPMVNHAV 327
Qy 301 LAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGKNMCGVATCASYPIVA 351
Db 328 LAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGKNMCGVATCASYPIVA 378

RESULT 10
US-10-087-714-22
; Sequence 22, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; APPLICANT: Dixon, Richard A.
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMCI0099
; CURRENT APPLICATION NUMBER: US/10/087,714
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 09/462,576
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US98/14895
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/052,604
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: 60/272,415
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; TYPE: PRT
; ORGANISM: Nicotiana tobaccum
US-10-087-714-22

Query Match 69.0%; Score 1286; DB 14; Length 360;
Best Local Similarity 69.6%; Pred. No. 5, 8e-127;
Matches 247; Conservative 35; Mismatches 65; Indels 8; Gaps 5;
Qy 5 LLEFLLFLVSALSVALAG---FEEDNPIRS-VTQRPDSIEPAILGVLSGCRHAFHAFPA 60
Db 6 LLLALVAGGLFASALAGPATADENPIRVQVSDGLHELENAILQVVGKTRHLSAFRFA 65
Qy 61 RYRGKSYGSEBEIKKRFGLFVENLAFIRSTNKRDLSTYLGINOPADLTWEEFRTNRLG 120
Db 66 HRIGKRYESVEIKRFEVFLDNLKKIRSHNKKGLSYKLVGNFEFDLTWEEFRTNRLG 125
Qy 121 QNCSTAHGNHREVDGVLVPTDRDREQGVSPVKDQSCGS-WTFSTTGALAAAYTQLTGS 179
Db 126 QNCSTATTGNLKVTVNLVPTDRDREAGIVSPVKQKQCGSCWTFTSTTGALAAAYSAQF 185
Qy 180 S--TLSEQQLVDCASAFNNFGC--GGLPSQAFVYKNGGIDTEQTPYLVGMGICNFKQ 236
Db 186 KGISLSEQQLVDCAGAFNFGCGGLPSQAFYIKYNGGLDTEEAIPYTKDGVCKFT 245
Qy 237 NVGVKVIDSINITLGADELKHAVALVRPVSVAFEVVKGFNLYKKGVSDDTCGRDPMV 296
Db 246 NVGVKVIDSVNITLGADELKHAVALVRPVSVAFEVVKGFNLYKKGVSDDTCGRDPMV 305
Qy 297 NHAVLAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGKNMCGVATCASYPIVA 351

Db 306 NHAVLAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGKNMCGVATCASYPIVA 360

RESULT 11
US-10-424-599-195993
; Sequence 195993, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 195993
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19007C.1.pep
US-10-424-599-195993

Query Match 68.6%; Score 1280; DB 12; Length 353;
Best Local Similarity 68.3%; Pred. No. 2, 4e-126;
Matches 244; Conservative 34; Mismatches 69; Indels 10; Gaps 5;
Qy 1 MAACKLFFLLFLVSALSVALAG--FEEDNPIRSVTQRPDSIEPAILGVLSGCRHAFHAF 58
Db 1 MARLSLLIFAFCAVAVAVAGSSFFDDANPRLAS---DLESQVLDDVIGQSRHLSAF 56
Qy 59 FARRYKSYGSEBEIKKRFGLFVENLAFIRSTNKRDLSTYLGINOPADLTWEEFRTNRLG 118
Db 57 FARRHGRYRSVDEIRNFRIFSDNLKLRSTNRRSLTYTLGVNHFADWTWEEFTRHKL 116
Qy 119 AAQNCSTAHGNHREVDGVLVPTDRDREQGVSPVKDQSCGS-WTFSTTGALAAAYTQL 177
Db 117 APQNCSTATLKGHRLTDAVLPEKDKWRKEGIVSQVQKQCGSCWTFTSTTGALAAAY 176
Qy 178 TGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFVYKNGGIDTEQTPYLVGMGICNFK 234
Db 177 FGKNIISLSEQQLVDCAGAFNFGCGGLPSQAFYIKYNGGLDTEEAIPYTKDGVCKFT 236
Qy 235 QENVGVKVIDSINITLGADELKHAVALVRPVSVAFEVVKGFNLYKKGVSDDTCGRDPM 294
Db 237 AKXAVRVVIDSINITLGADELKHAVALVRPVSVAFEVVKGFNLYKKGVSDDTCGRDPM 296
Qy 295 DVNHAVLAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGKNMCGVATCASYPIVA 351
Db 297 DVNHAVLAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGKNMCGVATCASYPIVA 353

RESULT 12
US-10-425-114-44426
; Sequence 44426, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128

```

; SEQ ID NO 44426
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700980869_FLI.pap
US-10-425-114-44426

Query Match      68.6%; Score 1280; DB 12; Length 353;
Best Local Similarity 68.3%; Pred. No. 2.4e-126;
Matches 244; Conservative 34; Mismatches 69; Indels 10; Gaps 5;

QY 1 MAKKLLPFLPLVLSALVALAG--FEEDNPIRSVTQRPDSEIPAILGVLSGCHAFHAF 58
DB 1 MARLSLLIFAFCAVAVAVAGSFFDDANPIRLAS-----DLSEOVLDVIGQSHALSFA 56

QY 59 FARRYGKSYGSEBIKKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFRNRLG 118
DB 57 FARRHGKRYRVDIRNFRIFSDNLKIRSTNRESLTTLGVNHFADLTWEEFRHKL 116

QY 119 AAOQCSATAHGNHRFVDGLPVTDRWREOGIVSPVKDQSGCS-WTFSTTTGALEAAATQ 177
DB 117 APQCSATLKGHNRLTDAVLDPEDKWRKEGIVSQVKDQSGCSWTFSTTTGALEAAATQ 176

QY 178 TGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVVKYNGGIDTEQTPYVLGYMGICNF 234
DB 177 FGKNIUSEQQLVDCAGAFNNFGCNGGLPSQAFYIKYNGGLDTEEAAYPTGKGVCKFT 236

QY 235 QENVGVKVIDSINITLGADELKHAVGLVRPVSVAFAEVKGNLYKKGVCATCASYPIVA 294
DB 237 AKNAVAVRVIDSINITLGADELKHAVGLVRPVSVAFAEVKGNLYKKGVCATCASYPIVA 296

QY 295 DVNHAVLAVGYGVEDGVPYWLKIKNSWGTNMGDNGYFKMELGKMGVCATCASYPIVA 351
DB 297 DVNHAVLAVGYGVEDGVPYWLKIKNSWGTNMGDNGYFKMELGKMGVCATCASYPIVA 353

RESULT 13
US-10-087-714-20
; Sequence 20, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMCI0099
; CURRENT APPLICATION NUMBER: US/10/087,714
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2000-05-22
; PRIOR FILING DATE: 1998-07-15
; PRIOR FILING DATE: 1997-07-15
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-087-714-20

Query Match      68.2%; Score 1272.5; DB 14; Length 360;
Best Local Similarity 66.4%; Pred. No. 1.6e-125;
Matches 239; Conservative 40; Mismatches 72; Indels 9; Gaps 4;

QY 1 MAKKLLF---FLLFLVLSALVALAGFEEDNPIRSVTQRPDSEIPAILGVLSGCHAFHFA 57
DB 1 MSAKTLSSWLVLVLAASAAANIGFDESNPIRMVSDGLREVEESVQLQSGRHVLSFA 60

; SEQ ID NO 44426
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700980869_FLI.pap
US-10-425-114-44426

Query Match      68.6%; Score 1280; DB 12; Length 353;
Best Local Similarity 68.3%; Pred. No. 2.4e-126;
Matches 244; Conservative 34; Mismatches 69; Indels 10; Gaps 5;

QY 1 MAKKLLPFLPLVLSALVALAG--FEEDNPIRSVTQRPDSEIPAILGVLSGCHAFHAF 58
DB 1 MARLSLLIFAFCAVAVAVAGSFFDDANPIRLAS-----DLSEOVLDVIGQSHALSFA 56

QY 59 FARRYGKSYGSEBIKKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFRNRLG 118
DB 57 FARRHGKRYRVDIRNFRIFSDNLKIRSTNRESLTTLGVNHFADLTWEEFRHKL 116

QY 119 AAOQCSATAHGNHRFVDGLPVTDRWREOGIVSPVKDQSGCS-WTFSTTTGALEAAATQ 177
DB 117 APQCSATLKGHNRLTDAVLDPEDKWRKEGIVSQVKDQSGCSWTFSTTTGALEAAATQ 176

QY 178 TGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVVKYNGGIDTEQTPYVLGYMGICNF 234
DB 177 FGKNIUSEQQLVDCAGAFNNFGCNGGLPSQAFYIKYNGGLDTEEAAYPTGKGVCKFT 236

QY 235 QENVGVKVIDSINITLGADELKHAVGLVRPVSVAFAEVKGNLYKKGVCATCASYPIVA 294
DB 237 AKNAVAVRVIDSINITLGADELKHAVGLVRPVSVAFAEVKGNLYKKGVCATCASYPIVA 296

QY 295 DVNHAVLAVGYGVEDGVPYWLKIKNSWGTNMGDNGYFKMELGKMGVCATCASYPIVA 351
DB 297 DVNHAVLAVGYGVEDGVPYWLKIKNSWGTNMGDNGYFKMELGKMGVCATCASYPIVA 353

RESULT 13
US-10-087-714-20
; Sequence 20, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMCI0099
; CURRENT APPLICATION NUMBER: US/10/087,714
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2000-05-22
; PRIOR FILING DATE: 1998-07-15
; PRIOR FILING DATE: 1997-07-15
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-087-714-20

Query Match      67.6%; Score 1261.5; DB 14; Length 362;
Best Local Similarity 66.3%; Pred. No. 2.3e-124;
Matches 240; Conservative 44; Mismatches 67; Indels 11; Gaps 6;

QY 1 MAKKLLFLLFL-----VSALSVALAGFEEDNPIRSVTQRPDS--IEPAILGVLSGCHAF 54
DB 1 MAHERILLJLAAVAATAVSAVAASGFDSDNPIRSVTDHAAASALESTVIALGRTRGAL 60

QY 55 HFARFARRYKSYGSEBIKKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFR 114
DB 61 RFARFARVGHKRYGDAAEVQRRFRIFSELSLELVRSTNRGLPYRLGINRPFADMSWEEFQA 120

QY 115 NRIAGAAQNCASATAHGNHRFVDG--VLPVTRDWRDEOGIVSPVKDQSGCS-WTFSTTTGALEA 172
DB 121 SRLGAAQNCASATLGNHRVRDAPALPETKDWRDGIIVSPVKDQSGCSWTFSTTTGSLEA 180

QY 173 AYTQLTGS--TJSEQQLVDCASAFNNFGC--GGLPSQAFYVVKYNGGIDTEQTPYVLGYMG 229
DB 181 RYVQATGPPVLSLSEQQLADCATRYNNFGSGGLPSQAFYIKNYNGGLDTEEAAYPTGVNG 240

QY 230 ICNFKQENVGVKVIDSINITLGADELKHAVGLVRPVSVAFAEVKGNLYKKGVCATCASYPIVA 289
DB 241 ICHYPENAGHKVLDVSNITLVADELKNAVGLVRPVSVAFAEVKGNLYKKGVCATCASYPIVA 300

```

Search completed: March 29, 2004, 22:28:33
Job time : 333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 22:17:04 ; Search time 3294 Seconds

(without alignments)

4631.678 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAAKLLFLLPLVLSALVAL.....ELGKMGVATCASYPIVAV 352

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USFTO_spool/US10087714/runat_26032004_113452_933/app.query.fasta_1.519
-DB=genEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10087714@cgn 1 1 2372 runat_26032004_113452_933 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rod:*
36: em.htg.man:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1322.5	70.9	1382	8	MZECYPA
2	1320.5	70.8	1486	8	AK066748
3	1310.5	70.3	1412	8	ZMSEEL
4	1310.5	70.3	1442	6	A43549
5	1300.5	69.7	1407	6	AR231148
6	1297	69.5	1332	8	AF454960
7	1286	69.0	1340	8	AB032168
8	1282	68.7	1345	8	LMU249847
9	1281.5	68.7	1468	8	RICOZC
10	1279.5	68.6	1305	8	AF233883
11	1279.5	68.6	1331	8	BT000676
12	1279.5	68.6	1352	8	AY088662
13	1279.5	68.6	1355	8	AF083703
14	1279.5	68.6	1370	8	AF360273
15	1279.5	68.3	1409	8	BT000673
16	1274.5	68.3	1366	8	BT000674
17	1265.5	67.9	1402	8	AY091771
18	1262.5	67.7	1347	8	PHU31094
19	1261	67.6	1444	8	PAU33166
20	1254.5	67.3	1381	8	PSRNACP
21	1252.5	67.2	1361	8	PSA278699
22	1249	67.0	1388	6	A43538
23	1249	67.0	1388	8	LECYP3GN
24	1204	64.6	1484	8	CACG2
25	1000	53.6	91274	8	ATF18N11
26	929	49.8	77298	8	AB011483
27	822.5	44.1	2966	8	AY017412
28	796	42.7	1362	10	RNCATH
29	792	42.5	1149	4	SSAF001169
30	783.5	42.0	937	8	AY383240
31	779	41.8	1360	10	RATCATH
32	765	41.0	1381	10	MCCTSH
33	763	40.9	972	9	AF426248
34	763	40.9	1008	9	AF426247
35	763	40.9	1372	10	BC006878
36	763	40.9	1399	9	HSCATH
37	759	40.7	1470	9	BC002479
38	757	40.6	1438	9	AK130158
39	744	39.9	14780	8	AP005419
40	722	38.7	536	8	AF213939
41	695	37.3	5372	8	HVLEU
42	691	37.1	1106	6	AX774741
43	691	37.1	1106	9	HSCATH
44	671	36.0	1410	8	AK109136
45	661.5	35.5	1168	3	AY336797

ALIGNMENTS

RESULT 1

MZECYPA 1382 bp mRNA linear PLN 10-FEB-1999
 LOCUS
 DEFINITION Corn mRNA for cysteine proteinase, clone CCP2, complete cds.
 ACCESSION D45403
 VERSION D45403.1 GI:1688044
 KEYWORDS cysteine proteinase.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
 Domoto, C., Watanabe, H., Abe, M., Abe, K. and Arai, S.
 Isolation and characterization of two distinct cDNA clones encoding corn seed cysteine proteinases
 JOURNAL Biochim. Biophys. Acta 1263 (3), 241-244 (1995)
 MEDLINE 96004895
 PUBMED 7548211
 REFERENCE 2 (bases 1 to 1382)
 AUTHORS Domoto, C.
 TITLE Direct Submission
 JOURNAL Submitted (31-JAN-1995) Chieko Domoto, Atomi Junior College; Ohtsuka 1-5-2, Bunkyo-ku, Tokyo 112, Japan
 (Tel:03-3941-8161(ex.575), Fax:03-3945-1036)
 COMMENT On Nov 27, 1996 this sequence version replaced gi:644489.
 Sequence updated (26-Oct-1996) by: Chieko Domoto.

FEATURES

Location/Qualifiers
 1..1382
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="taxon:4577"
 /clone="CCP2"
 52..1134
 /codon_start=1
 /product="cysteine proteinase"
 /protein_id="BA008245.1"
 /db_xref="GI:164490"
 /translation="MPPRRFLVAVLVVLAADTAAVVNSGFADSNPIRPVTDRAALES
 TVFAALGRDALARFARVRYGKSYESAEEVHVRIFSESILQVRSNRKGLSVRL
 GINRFADMSHEPRALGNAONCSATLGNHMRAAVALPETKDWREDGLSPVFN
 QHCGCSFTSTTGALAAATONATKPSLSEQOLVDCGFANFGCNGGLPSQAFY
 IKYNGGLDTEESTPYGVNGVICFKFKNVGVKVDVSNITLGADELKDAVGLRFPVS
 VAPEVITGFRLYKSGVYTSDDCTTPMDVNHVAVLVGVGVEDVGPVWLLIKNSWGADWG
 DEGYFKMEMGKMGVATCASYPIVA"

CDS

ORIGIN

Alignment Scores:
 Pred. No.: 7,49e-111 Length: 1382
 Score: 1322.50 Matches: 258
 Percent Similarity: 79.44% Conservative: 28
 Best Local Similarity: 71.67% Mismatches: 65
 Query Match: 70.91% Indels: 9
 DB: 8 Gaps: 6

US-10-087-714-2 (1-352) x MZECYPA (1-1382)

Qy 1 MetAlaAlaLysLeuLeuPhePheLeuPheLeuValSerAlaLeuSerValAlaLeu 20
 Db 52 ATGGTCCACCGCGCTGTTCCTCGCGCTGCTCTCTCGCGGACACCGCGCGCTG 111
 Qy 21 -----AlaGlyPheGluGluAspAspProIleArgSerValThrGlnArgProAspSer 38
 Db 112 GTCAACTCCGGCTTCGGCGGCTCCACACCGGATCCCGCGCTCAGCGACCGCGCGCTCC 171
 Qy 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
 Db 172 CGCTCGAGTCCACGGTCTTCGCGCGCTTCGCGCGACCGCGCGCGCTCGCTCGCA 231
 Qy 58 ArgPheAlaArgArgTyrGlySerTyrGlySerGluGluIleLysLysArgPhe 77
 Db 232 CGCTTCGCGCGTACGATCGGCAAGAGCTACGAGAGCGCGCGGAGGTCACACGCGGTT 291
 Qy 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97

Db 292 AGGATCTTCTCCGAGAGCTCCAGCTGTCCTCGCTCCACCAACCGCAAGCGCTCTCTCTAC 351
 Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
 Db 352 CGCTTCGGCATCAACCGCTTCGCGGACATGATGCTGGGAGGAGTTCCTGCGACCGCGCTC 411
 Qy 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
 Db 412 GGCGACGCCAGAACTGCTCCGCCACCTTACCGGCAACCAACCGGATGCGCGCGCGCGCC 471
 Qy 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAsp 155
 Db 472 GTTCGGCTGCGGAGACGAAGACTGCGGGAGGATGGGATGTGACGCCAGTGAAAC 531
 Qy 156 GlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174
 Db 532 CAGGCGCACTGTGGATCATGTGGACCTTACGACTACTGTGGTGGTGGTGGTGGTGGTGGT 591
 Qy 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
 Db 592 ACCAGGCACTGGCAAGCCCATCTCTCTCTGAGCAACAGCTTGTGACTGTGGTGGTGGT 651
 Qy 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
 Db 652 GCATTCAACAATTTGGGATGCAACGAGGCGCTTCCATCCAGGCGCTTGTGAATACATCAA 711
 Qy 212 TyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCys 231
 Db 712 TACATGTGGCTTGCCTGACCTGAGGATCTTACCTTACCAAGGTGTCAATGGAATCTGC 771
 Qy 232 AsnPhelysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGly 251
 Db 772 AAGTTTAAGAAATGAGATGTTGGAGTCAAGGTTTGGACTCGTAAACATCACCTCGGT 831
 Qy 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 Db 832 GCTGAGGATGACGTAAGGATGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 891
 Qy 272 ValValLysGlyPheAsnLeuTyrIleLysGlyValTyrSerSerAspThrCysGlyArg 291
 Db 892 GTGATCACTGGTTTCAGGCTGTACAGAGCGGAGTTTACACTAGCGACCATTTGGAACT 951
 Qy 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311
 Db 952 ACACCGATGGATGTGAACCGCTGTTCTGGCTGTGGCTACGGTGTGGAAGTGGTGA 1011
 Qy 312 ProTyrTrpLeuLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
 Db 1012 CCTTACTGGCTCATCAAGAACTCATGGGCGCTGACTGGGCTGATGAGGGTTACTTCAAG 1071
 Qy 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
 Db 1072 ATGGAATGGGCAAGAACATGTGGCTGTGTCTAGTGTGATCTACCTACCTATTTCGCA 1131

RESULT 2

AK066748 1486 bp mRNA linear PLN 24-JUL-2003
 LOCUS Oryza sativa (japonica cultivar-group) cDNA clone J013074D19, full insert sequence.
 DEFINITION
 ACCESSION AK066748
 VERSION AK066748.1 GI:32976766
 KEYWORDS FLI CDNA; CAP trapper.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.

REFERENCE

1
 The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-Length cDNA Project Team., Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,

Db 844 CTGGGTGCTGAGGATGAGTGAAGATGCTGTGGACCTGTTCCTCAGTTAGTGTGCC 903
 QY 270 PheGluValValIsclyPheAsnLeuTyrlslylsyGlyValTySerSerAspThrCys 289
 Db 904 TTTCAGGTGATCAACGGTTTCAGGATGACAGAGTGGAGTTTACACAAGTACCATTGT 963
 QY 290 GlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrlslylsyGluAsp 309
 Db 964 GGAACCTTCTCCCAATGGATGATGAACACCGCTGTTCTGGCCGTGGCTATGGTGCMAAT 1023
 QY 310 GlyIleProTyTrpIleuIleAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyTr 329
 Db 1024 GCGGTCCCTACTGCTCATCAGAACTCATGGGTGCGAGCTGGGTGCAATGGTTAC 1083
 QY 330 PheIysMetGluLeuGlyIysAsnMetCysGlyValAlaThrCysAlaSerTyProIle 349
 Db 1084 TTCAAGATGGAATGGCAAGAACATGTGCGGTATTGCTACTTGGCATCTACCCCTATT 1143
 QY 350 ValAla 351
 Db 1144 GTTGCA 1149

RESULT 3

ZMSEB1
 LOCUS 1412 bp mRNA linear PLN 10-SEP-1997
 DEFINITION Z.mays mRNA for cysteine proteinase, Seel.
 ACCESSION X99336
 VERSION X99336.1 GI:1491773
 KEYWORDS cysteine proteinase; seel gene; senescence-related protein.
 SOURCE Zea mays
 ORGANISM Zea mays

REFERENCE

AUTHORS Griffiths C.M., Hosken S.E., Oliver D., Chojacki, J. and Thomas, H.
 TITLE Sequencing, expression pattern and RFLP mapping of a senescence-enhanced cDNA from Zea mays with high homology to oryzain gamma and aleurain
 JOURNAL Plant Mol. Biol. 34 (5), 815-821 (1997)
 MEDLINE 97422404
 PUBMED 9748172
 REFERENCE Griffiths C.M.
 AUTHORS Griffiths C.M.
 TITLE Direct Submission
 JOURNAL Submitted (08-AUG-1996) C.M. Griffiths, Inst. of Grassland and Environmental, Research, Plas Gogerddan, Aberystwyth, Ceredigion, SY23 3BB, UK

FEATURES

source

1..1412
 /organism="Zea mays"
 /mol_type="mRNA"
 /cui="taxon:ES"
 /db_xref="taxon:4577"
 /tissue_type="leaf"
 /clone_lib="lambda gt10"
 /dev_stage="senescing leaves"
 1..1412
 /gene="seel"
 63..1145
 /gene="seel"
 /codon_start=1
 /product="cysteine protease"
 /protein_id="CAA68192.1"
 /db_xref="GI:1491774"
 /db_xref="GOA:Q43705"
 /db_xref="SPTREMBL:Q43705"
 Translation="MAPRLLVAVVLAATAAANGSGADSNPIRPVTVDRASALESTVFAALGRTDALPFAFVAVYKSVESAFAVHKRPISESILQVRSNTRKGLSVRLGINFADMSWEFPATRLGAQNCASATLTGNHRAAVALPEIKMDREDGLVSPVKNCGHGSCTSTGALNAATGATGATPISUSEQVLVDICLAFNNFCNGGLPSQAFYIKYNGGLDTEESYPQGVNGISKFNENGVKVLDSVNTLGAEDLKDAGLVLRPS

gene

CDS

sig_peptide 147..164
 /gene="seel"
 /note="vacuolar sorting sequence"
 mat_peptide 489..1142
 /gene="seel"
 /product="unnamed"
 misc_feature 63..488
 /gene="seel"
 /note="pre-pro protein"

ORIGIN

Alignment Scores:
 Pred. No.: 9.Sle-110 Length: 1412
 Score: 1310.50 Matches: 257
 Percent Similarity: 78.61% Conservative: 26
 Best Local Similarity: 71.39% Mismatches: 68
 Query Match: 70.27% Indels: 9
 DB: 8 Gaps: 6
 US-10-087-714-2 (1-352) x ZMSEB1 (1-1412)
 QY 1 MetAlaAlaIysLeuLeuPhePheLeuValSerAlaLeuSerValAlaLeu 20
 Db 63 ATGGCCACCGCCCTGCTGCTCGCGCTCGCGCCCTCGCGCCACCGCGCGCG 122
 QY 21 Ala-----GlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAspSer 38
 Db 123 GCCAATCCGCGCTTCGCGGACTCCACCCGATCCGCGCGCTCACCACCGCGCGCTCC 182
 QY 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
 Db 183 CGCGTCGAGTCCACCGCTTCGCGCGCTCGCGCGCCACCGCGCGCTCGCGCTTCGCA 242
 QY 58 ArgPheAlaArgArgTyrlslylsySerTyrlslylsyGluGluGluIleIleIleIleIle 77
 Db 243 CGCTTCGCGCTCAGTACGCGAAGAGCTACGAGAGCGCGCGCGAGTCCACAAAGCGGTC 302
 QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgIysAspLeuSerTyTr 97
 Db 303 AGGATCTTCTCCGAGAGCTCCAGTGGTCCGCTCCACCAACCGCAAGCGCTCTCTAC 362
 QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
 Db 363 CGCTTCGCGCTCAACCGCTTCGCGGACATGAGCTGGAGAGTTCCTCGGACCGCGCTC 422
 QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
 Db 423 GGTGCGAGCCAGAACTGCTCCGCGCAGCTTACCGCGCAACCCACCGGATCGCGCGCGCC 482
 QY 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValIysAsp 155
 Db 483 GTTCGCTCGCGAGAGAGAAAGACTGGAGAGAGATGGGATTTGTAGCCAGTGAAGAAC 542
 QY 156 GlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyTr 174
 Db 543 CAGGCGCACTGTGGATCATGCTGGACCTTCAGCACTACTTGGTGCATTCGAGCTGCATAT 602
 QY 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
 Db 603 ACCGAGCAACTGGCAAGCCCATCTCTCTCTGAGCAACACGCTTGTACTGTGGTTG 662
 QY 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyTrValIys 211
 Db 663 GCATTCAACAAATTTTCGATGCAACGAGGCGCTTCCATCCCGCGCTTTGAATACATCAA 722
 QY 212 TyrAsnGlyGlyIleAspThrGluGlnThrTyTrProTyTrLeuGlyValMetGlyIleCys 231
 Db 723 TACATGGTGGCTTGGACTGAGACTGAGGAATCTTACCCCTTACCAAGGTGTCAATGGAAATCTCC 782
 QY 232 AsnPhelysGlnGluAsnValGlyValIysValIleAspSerIleAsnIleThrLeuGly 251

VAFVITGFRLYKSGVYTSDHCGTTPMDVNHVAVLAVGYGVYVEDGVYWLKNSWGDWG
 DEGYFKMEMGKMGVATCASYPIVA"


```

Db      68 CTGCTGATTCATCCGCGCATCGACGGCGGAGAGATATCGGATTGTGATGACGCAACCCG 127
Qy      29 lleAgservAlThrGlnArgProAspSerlleGluProAlaIleLeuGlyValIleuGly 48
Db      128 ATCCGTATGCTCTCGACCGTCTCCGGAGGTAGAGGAATCGTGTGTCACGATCTTAGC 187
Qy      49 SerCysArgHisAlaPheHisPheAlaArgPheAlaArgAgtTyrGlyLysSerTyrGly 68
Db      188 CAATCAGCCAGGTATCTCTTCCTCGCTCGCTCGCTCAGTACAGTACGGAAGAGTACGAG 247
Qy      59 SerGluGluGluIleLysLysArgPheGlyIlePheValGluIleAsnLeuAlaPheIleArg 88
Db      248 AACCGGAGGAGATGAAGCTCAGGTCTCGATATTCAGAGGAATCTGGATTGATCAGA 307
Qy      89 SerThrAsnArgLysAspLysSerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThr 108
Db      308 TCCACCAATAGAAGGCTTGTCTTACAACTCGGTGTCAATCAGTTTGTGTATGATGAGC 367
Qy      109 TrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHis 128
Db      368 TGGCAAGAGTTTCAAAGGACCAAGCTTGTGTCTGCTCAAACTGCTCTGCCACTTTAAAG 427
Qy      129 GlyAsnHisArgPheValAspGlyValIleuProValThrArgAspTyrArgGluGly 148
Db      428 GGAACCCCAAACTCACAGGAGAGCTCTTCGGAAACCAAGACTGGAGAGAGATGGT 487
Qy      149 IleValSerProValLysAspGlnGlySerCysGlySer---TrpThrPheSerThrThr 167
Db      488 ATTGTAGCCCTGTCAAAGACAGGGGGTGTGTGATCGTGTGACATTCAGTACTACT 547
Qy      168 GlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----ThrLeuSerGluGln 185
Db      548 GGAGCGCTTGGAGCGAGCTTATCACCAAGCATTTGGAAAGGAATATCACTCTCTGGACAA 607
Qy      186 GlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSer 204
Db      608 CAGCTTGTGATTTGTCTGGAGCTTCAACAACTATGCTGCAATGCTGGCTTCTCTCC 667
Qy      205 GlnAlaPheGluTyrValLysTyrAsnGlyIleAspThrGluGlnThrTyrProTyr 224
Db      668 CAAGCTTTGAATACATCAATCAACCGCGGCTCGACAGCGGAGGCTTATCTCTTAC 727
Qy      225 LeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyValLysValIleAsp 244
Db      728 ACCGGTGAAGATGGAACTGCAAGTATCTCTGAAACGTCGGTGTAGAGTCTCTTGAC 787
Qy      245 SerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArg 264
Db      788 TCAGTCAACATTAATCTCTGGTGTCTGAAGTGAATGAACTGAAGCATGCGGTGGATTGGTACGA 847
Qy      265 ProValSerValAlaPheGluValLysGlyPheAsnLeuTyrLysGlyValTyr 284
Db      848 CCAGTACGATGACATTCGAGGTATACACTCGCTTCGGCTTTACAGAGCGGAGTTTAC 907
Qy      285 SerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGly 304
Db      908 AGCGATAGTCACTGTGGACAGACTCCAATGACGCTGAACCACTGCTGTTTGGCGGTGGT 967
Qy      305 TyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrp 324
Db      968 TATGGAATTGAAGACGGTGTCCATATTGGCTTTATAAGAACTCATCGGGAGCAGATTGG 1027
Qy      325 GlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCys 344
Db      1028 GGTGACAAGGTACTTCAAGATGGAGATGGGAGAACATGTGTGTATTCGACATGT 1087
Qy      345 AlaSerTyrProIleValAla 351
Db      1088 GCATGTAACCGGTTGTGGCC 1108

```

RESULT 7
AB032168
LOCUS

AB032168 1340 bp mRNA linear PLN 24-JAN-2001

DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM

Nicotiana tabacum NTCP-23 mRNA for cysteine protease, complete cds.
AB032168
AB032168.1 GI:8347419
cysteine protease.
Nicotiana tabacum (common tobacco)
Nicotiana tabacum

REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiales; Solanales; Solanaceae; Nicotiana.
1 (sites)
Ueda,T., Seo,S., Ohashi,Y. and Hashimoto,J.
Circadian and senescence-enhanced expression of a tobacco cysteine protease gene
Plant Mol. Biol. 44 (5), 649-657 (2000)
21039173
11198425
2 (bases 1 to 1340)
Ueda,T.
Direct Submission
Submitted (03-SEP-1999) Tadamasu Ueda, National Institute of Agricultural Resources, Department of Molecular Genetics; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:uechue@affrc.go.jp, Tel:81-297-38-7443, Fax:81-298-38-7044)
Location/Qualifiers
1..1340
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Sameun NN"
/db_xref="taxon:4097"
/clone="NTCP-23"
/tissue_type="leaf"
/clone_lib="Lambda ZAP healthy tobacco"
/tissue_lib="healthy leaf"
1..1340
/gene="NTCP-23"
35..1117
/gene="NTCP-23"
/codon_start=1
/product="cysteine protease"
/protein_id="BAA9501.1"
/db_xref="GI:8347420"
/translation="MSRPSLLALVVAGLPSALAGPATADENPIROVSDGLHEL
ENALIQVCKTRHALSPARFAHYKRYESVEIEIKQFEVFLDNLKMRSHNKGILSY
KIGNEFTLTWDFRRDRQAQNCSTATTGNLKVTVLVLPETKDMRAGLIVSPYKN
QKQSCSWTFTTGALEAYSGAKGSLSQQLVDCAGAFNFCNGGLPSQAFY
IKSNGLDTEAYPTGKNGLCFSENVGVKVIDSVNITLGADELKVALVRPS
IAFEVKGFKYKSGVYSTECCNTPMVDVNHAVLAVGVGVENGVPYWLKNSWGDWG
DNGYFKWEMGKMKCGIATCASPVA"
125..142
/gene="NTCP-23"
/note="vacuolar sorting sequence"
461..1114
/gene="NTCP-23"
/product="cysteine protease"
/function="protein degradation"
/note="putative"

FEATURES source

1..1340
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Sameun NN"
/db_xref="taxon:4097"
/clone="NTCP-23"
/tissue_type="leaf"
/clone_lib="Lambda ZAP healthy tobacco"
/tissue_lib="healthy leaf"
1..1340
/gene="NTCP-23"
35..1117
/gene="NTCP-23"
/codon_start=1
/product="cysteine protease"
/protein_id="BAA9501.1"
/db_xref="GI:8347420"
/translation="MSRPSLLALVVAGLPSALAGPATADENPIROVSDGLHEL
ENALIQVCKTRHALSPARFAHYKRYESVEIEIKQFEVFLDNLKMRSHNKGILSY
KIGNEFTLTWDFRRDRQAQNCSTATTGNLKVTVLVLPETKDMRAGLIVSPYKN
QKQSCSWTFTTGALEAYSGAKGSLSQQLVDCAGAFNFCNGGLPSQAFY
IKSNGLDTEAYPTGKNGLCFSENVGVKVIDSVNITLGADELKVALVRPS
IAFEVKGFKYKSGVYSTECCNTPMVDVNHAVLAVGVGVENGVPYWLKNSWGDWG
DNGYFKWEMGKMKCGIATCASPVA"
125..142
/gene="NTCP-23"
/note="vacuolar sorting sequence"
461..1114
/gene="NTCP-23"
/product="cysteine protease"
/function="protein degradation"
/note="putative"

sig_peptide mat_peptide

1..1340
/gene="NTCP-23"
/note="vacuolar sorting sequence"
461..1114
/gene="NTCP-23"
/product="cysteine protease"
/function="protein degradation"
/note="putative"

Alignment Scores: Pred. No.: Score: Percent Similarity: Best Local Similarity: Query Match: DB:

1..51e-107
1286.00
79.72%
69.30%
68.95%
8
Matches:
Conservative:
Mismatch:
Indels:
Gaps:
1340
246
37
64
8
5

US-10-087-714-2 (1-352) x AB032168 (1-1340)

Qy 5 LeuLeuPhePheLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGly----- 22
Db 50 CTCCTATTGGCTCTGCTGCTCGCGGTGGCTTTTTCGCTCGCACTCGCCGACCGCGG 109

Qy	23	--PheGluAspAsnProIleArgSer---ValThrClnArgProaspSerIleGlu	40
Dd	110	ACCTTTGCCAGTAATCCGATCGATCAACAAGTCGTTCCTCACGGTATTACATGACGCAGGAG	169
Qy	41	ProAlaIleLeuGlyValLeuGlySerCysAargHieAlaPheHisPheAlaAargPheAla	60
Dd	170	AACGCAATTCTCCAAGTCGTTCGSCAACACCGCATCTCTCTCTCGCTCGCTTGCT	229
Qy	61	ArgArgTyrgLylysSerTygLyserGluGluGluLilelylsysAargPheGlyllePhe	80
Dd	230	CACAGGTATGGGAAGAGGTACAGATCAGTTGAGGAGATAAGAACAAGGTTTCGAGGTATTT	289
Qy	81	ValGluAsnLeuAlaPheIleArgSerThrAsnAargLyasPseusertyrThrLeuGly	100
Dd	290	TTGACAATTGGAAGATGATTTCGATCGACCAACAAGAAAGACTATCATCAAACACGGT	349
Qy	101	IleAsnGlnPheAlaaspLeuthTrtPglUglUpheArgThrAsnAargLeuGlyAlaAla	120
Dd	350	GTCATGAGTTTACCAGACCTAACATGGGACGAGTTCGGAGACAGACTGTGGGGCGAGCT	409
Qy	121	GlnAsnCysSerAlaThrAlaHiSclYasnHiAargPheValAaspGlyValLeuProVal	140
Dd	410	CAAACTGTTTCAGCCACCAAAAAGGGCAATCTCAAAGTCACTAACTGTTTCTGCCGGAG	469
Qy	141	ThrAargaspTrpAsgGluGlnGlylIeValSerProVallysaspGlnGlySerCysGly	160
Dd	470	ACGAAGACATGGCGGAGCTGGGATTTGTTCAGCCCACGTCGAAGAACCAAGGCGAATGCGGA	529
Qy	161	Ser---TrpThrPheSerThrThrGlyAlaLeuGlualalaatyrrThrGlnLeuThrGly	179
Dd	530	TCITGCTGGCAATTCAGCACACTACTGGTGCACTAGAAGCAGCATATATCCAAAGCATTTGGG	589
Qy	180	Ser-----ThrLeuSerGluGlnGlnLeuValaspCysalaserAlaPheAsnAsnPhe	197
Dd	590	AAGGGAATCTCTCTATCTGAGCAGCAGCTGTGTGACTGTGCGAGCTTTTAATAACTTT	649
Qy	198	GlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrrVallystyrAsnglyVglyle	216
Dd	650	GGCTGCAATGGTGGCTCCCCATCACAGCCITTTCAGTATATAAATCCAATGGTGTCCTT	709
Qy	217	AspThrGluGlnThrTyrrProtyrLeuGlyValMetGlylleCysAsnPheLysGlnGlu	236
Dd	710	GACACTGAAGAAGCATATCCATACACTGGCAAGAAATGGCTTATGTAATTTCTCATCAGAA	769
Qy	237	AsnValGlyVallysValIleAaspSerIleAsnIleThrLeuGlyValAgLuAaspGluLeu	256
Dd	770	AAETGTGTGTCGCAAGTCATCGATTCGTGCAATATTACCTCGGTGCTGAGATGAAC TA	829
Qy	257	LysHisAlavalGlyLeuValArspovalSerVallapheGluValvalLysGlyPhe	276
Dd	830	AAAAACGCGGTTCGATTTGGTTAGGCCCGCTTAGTAGCTTTWAGGTGTATAAAGGTTTC	889
Qy	277	AsnLeutyryLysLysGlyValTyrrSerSerAspThrCysGlyAyGaspProMetAspVal	296
Dd	890	AAACNATACAGAGTGGTGTGTTTTACACAGCACCGNAATGGCGCAACATCCCATGATGTA	949
Qy	297	AsnHisAlavalLeuAlaValGlyTyrrGlyValGluAaspGlylleProTyrrTrpLeulle	316
Dd	950	AACCATGCTGTTCTTGCTGTGGGTACGCGTGTGAAAATGTGTGTCCCTATTGCTCATC	1009
Qy	317	LysAsnSerTrpGlyThrAsnTrpGlnVAspAsnGlyTyrrPheLysMetGluLeuGlyLys	336
Dd	1010	PAGAATTTCATGGGAGCAGATTGGGTGCACATGGATACTTCAAAATGGAGATCGGAAG	1069
Qy	337	AsnMetCysGlyValAlaIathrcysAlaserTyrrProfileValAla	351
Dd	1070	AACATGTGTGTATTGCCACTTTGGCATCTCCACCTGATCGCTGGCTGGC 1114	

AJ249847.1 GI:6851029
 cysteine protease; seel gene.
 Lolium multiflorum (Italian ryegrass)
 Lolium multiflorum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Poae; Lolium.
 1
 Li, Q., Bettany, A. J., Donnison, I., Griffiths, C. M., Thomas, H. and
 Scott, I. M.
 Characterisation of a cysteine protease cDNA from Lolium
 multiflorum leaves and its expression during senescence and
 cytokinin treatment
 Biochim. Biophys. Acta 1492 (1), 233-236 (2000)
 20461783
 PUBMED 11004495
 2 (bases 1 to 1345)
 Li, Q.
 Direct Submission
 Submitted (30-SEP-1999) Li Q., Department of cell Biology,
 Institute of Grassland and Environmental Research, Plas Gogerddan,
 Aberystwyth, Ceredigion, SY23 3BB, UNITED KINGDOM
 Location/Qualifiers
 1..1345
 /organism="Lolium multiflorum"
 /mol_type="mRNA"
 /db_xref="taxon:4521"
 1..1345
 /gene="seel"
 36..1115
 /gene="seel"
 /function="essential for proteolysis"
 /codon_start=1
 /product="cysteine protease"
 /protein_id="CAB71032.1"
 /db_xref="GI:6851030"
 /db_xref="GOA:Q9M4E5"
 /db_xref="SPRTEMBL:Q9M4E5"
 /translation="MAKGRITLVLAVALATAVAASAFDSNPRTPTERAASAVES"
 VLGALGTRIALPAPFAVRPHGKSYGSAAEVQRFPRFBSLDEVRSTNRKLSYKLG
 INRSDTWTFEFGALQAUAQTGATLGNHLMRDNALPETHKWRGTGIVSPVYDQD
 SCGSWTFSTFQALAEAYATQATGKNIISBQLVDCAGAVNFCNGGLDPSQPEYIN
 YNGIDTSESYGVNGVCKYKRPENAAVQAVDSVITLNAEDSLKNAVLGLVRPVSVA
 FEVIDGPKYKGYVTSDDGTTDPDYNHVALVAVGYGVNGVPTVLLIKNSWGDWMDGE
 GYKFWKMGKMCVATCATCASYPIILAA"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,51e-107 Length: 1345
 Score: 1282.00 Matches: 243
 Percent Similarity: 80.51% Conservative: 42
 Best Local Similarity: 68.64% Mismatches: 63
 Query Match: 68.74% Indels: 6
 DB: 8 Gaps: 5
 US-10-087-714-2 (1-352) x LMU249847 (1-1345)
 Qy 4 LysLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPhe 23
 Db 48 CGCATCTCTGCTTGGCGCTCGCGTCTCTGSCACCGCGCGAGTGGCGCTTCCTTT 107
 Qy 24 GluGluAspAenProIleArgSerValThrGlnArgProAspSer---IleGluProAla 42
 Db 108 GCCGACTCCAAACCGGATCCGCCCGCTCACCGAGCGCGCGCTCGGGGTGCGATCAACG 167
 Qy 43 IleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgArg 62
 Db 168 GTCTCTGGCGGCTCTGGCGGACCGGACACCGCTTCGCTTCGGCGGCTTCGCGTCAAG 227
 Qy 63 TyrGlyIysSerTyrGlySerGluGluGluIleIysIysArgPheGlyIlePheValGlu 82
 Db 228 CACGGCAAGACTACGGAGGCGCGCGGAAGTGCAAAGCGCGGTTCGGGATCTTTCGAG 287


```

Qy 115 AsnArgLeuGlyAlaAlaGlnAsnCySerAlaThrAlaHisGlyAenHisArgPheVal 134
Db 407 AGCCGGTTCGGCGGGCGGAGAACTGCTCGGAGAGCTCCCGGCAACCAACCGGATGCCG 466
Qy 135 AspGly---ValLeuProValThrArgAspTrpArgGluGlnGlyLeValSerProVal 153
Db 467 GACCGCGCCGCTCTCCCGGAGACCAAGACTGGAGGAGGATGGATCGTGAGCCCGTG 526
Qy 154 LysAspGlnGlySerCySgLySer---TrpThrPheSerThrThrGlyAlaLeuGluAla 172
Db 527 AAAGACAGGGTCACTGGCTCTGTTGGCCCTTCAGACCACTGGCTCTCTTGAGGCA 586
Qy 173 AlaTyThrGlnLeuThrGlySer-----ThrLeuSerGluGlnLeuValAspCys 190
Db 587 CGATATACTCAGGCCACTGGGCCCTGTCTCTCTTTCTGAGCAGCAGCTGGCTGATTGT 646
Qy 191 AlaSerAlaPheAenAsnPhelGlyCys---GlyGlyLeuProSerGlnAlaPheGluTy 209
Db 647 GCTACTCGATACAATAATTCGGATGAGTGAGGAGGCTTACATCTCAGGCCCTTTGAGTAC 706
Qy 210 VallysTyThrAsnGlyGlyIleAspThrGluGlnThrTyProTyThrLeuGlyValMetGly 229
Db 707 ATCAATACAAATGCTGGCCCTTGACACTGAAGAGGCTTACCCTTACACGGGTGTCAATGGC 766
Qy 230 IleCysAenPhelGlyGlnGluAsnValGlyValIleValIleAspSerIleAenIleThr 249
Db 767 ATCTGTCTATACAAAGCCCTGAAACCGCTGGAGTCAAGCTTTTGGACTCCGTAAACATCACC 826
Qy 250 LeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAla 269
Db 827 CTGCTTCTGAGGATGAGCTGAAGATGCTGTTGGACTTGTCTGTCAGTTAGTGTGCC 886
Qy 270 PheGluValValLysGlyPheAenLeuTyriLysGlyValTySerSerAspThrCys 289
Db 887 TTTGAGTGTATCAACGGCTTTCAGGATGATCAAGAGTGGAGTTTACACAAAGTACCATTGT 946
Qy 290 GlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyriGlyValGluAsp 309
Db 947 GGAATCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006
Qy 310 GlyIleProTyThrLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTy 329
Db 1007 GGGCTTCCCTACTCGCTCATCAAGAACTCATGGGGTGCAGACTGGGGTGCAGAAATGTTAC 1066
Qy 330 PheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyProIle 349
Db 1067 TTCAGATGGAATGGGCAAGAACTGTCGGGTATTGCTACTTCTGCGCATCTACCCATT 1126
Qy 350 ValAla 351
Db 1127 GTTGCA 1132

RESULT 10
LOCUS AF233883
DEFINITION Arabidopsis thaliana AALP protein mRNA, complete cds.
ACCESSION AF233883
VERSION AF233883.1 GI:7230639
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 1305)
AUTHORS Masuoka, K. and Raikhel, N.V.
TITLE The plant vacuolar sorting receptor ATELP is involved in transport of NH(2)-terminal propeptide-containing vacuolar proteins in Arabidopsis thaliana
JOURNAL J. Cell Biol. 149 (7), 1335-1344 (2000)
MEDLINE 20330156

```

```

10871276
2 (bases 1 to 1305)
Raikhel, N.
Direct Submission
Submitted (13-FEB-2000) MSU-DOE Plant Research Lab, MSU, Plant
Biology Building, East Lansing, MI 48824, USA
Location/Qualifiers
1. 1305
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
10. 1086
/note="similar to aleurain"
/codon_start=1
/product="AALP protein"
/protein_id="AAF43041.1"
/db_xref="GI:7230640"
/translaton="MSAKTILSSVVLVLAASAAANIGFDESINPIRMVSDGLREVVE
SVSILGOSRHVLSFARTRHYKKYQNVEMKLFISIFKENLDLIRSTNKKGLSYKL
SVSILGOSRHVLSFARTRHYKKYQNVEMKLFISIFKENLDLIRSTNKKGLSYKL
GCGSCADLTWQEFORTKLAONCSATLKSHKYTEALPETKDRSDGI VSPVKDQG
GCGSCADLTWQEFORTKLAONCSATLKSHKYTEALPETKDRSDGI VSPVKDQG
SNGGLDTEKAYPTTKDETKFSAENVGVQVLSNVITLGADEBKAHAGLVRRPVSI
FEVTHSELYKSGYVTTSDHCGSTPMDVNHAVLAVGYGVEDGVPYWLINSGWADWGD
GYFKMEMKMKMGCIATCASYPVVA"
ORIGIN
Alignment Scores: 5.7e-107 Length: 1305
Pred. No.: 1279.50 Matches: 240
Score: 77.93% Conservative: 39
Percent Similarity: 77.93% Mismatches: 72
Best Local Similarity: 67.04% Indels: 7
Query Match: 68.61% Gaps: 4
DB: 8

US-10-087-714-2 (1-352) x AF233883 (1-1305)
Qy 1 MetalAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 10 ATGTCGCGAACAATCTATCATCATGAGTATTTGGTGTCTCGTCGCGCATCAGCA 69
Qy 18 ValAlaLeuAlaGlyPheGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 70 GCGGCTAATATCGGATCGATGAGTCAACCCGATCCGATCGTCTCGATGGTCTCCGG 129
Qy 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 130 GAGGTAGAGATCTGTTCCAGATCTTAGTCAATCTCGTCACTCTCTCTCTCGCT 189
Qy 58 ArgPheAlaArgArgTyriGlyLysSerTyriGlySerGluGluGluIleLysLysArgPhe 77
Db 190 CGCTTCACTCACCGATATGTTAAAGATATCAGAACGTCGAGGAGATGAAGCTTCGATT 249
Qy 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTy 97
Db 250 TCGATTTTCAGGAGAACTTTGATTGATCAGATCCACCAACAAAGAGGCTTATCTTAC 309
Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 310 AAACCTCGGTGTTAATCAATTTGCTGATTGATGTCGCAAGAGTTTCAAGAACCAAGCTT 369
Qy 118 GlyValAlaGlnAsnCySerAlaThrAlaHisGlyAenHisArgPheValAspGlyVal 137
Db 370 GTTGCTGCTCAGAACTGCTTGCCACTTAAAGGCGCAGCCACAGGTTCACAGAACGCT 429
Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyLeValSerProValLysAspGlnGly 157
Db 430 CTTCTGAAACAAAGACTGGAGAGAGATGATGATCTGTTAGTCCCGTCAAGATCAGGGA 489
Qy 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyThrGln 176
Db 490 GTTGTGGATCTTGTGACATTCAGCAACTGAGGCTCTTGAGCGAGCTTACCATCAG 549
Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194

```



```

Db      550 GCATTTGGAAAAGGATATCTCTCTGAGCAACAGCTTGTGGATTGCTGAGCTTTC 609
QY      195 AsnAsnPhcGlyCys--GlyGlyLeuProSerGlnAlaPheGluTyVallystYrAsn 213
Db      610 ATAATACTAGTGTGCAATGGTGGCCCTCTCTCAGCCCTTTGAATACATAAATCCAAAC 669
QY      214 GlyGlyLeuAsnThrGluGlnThrTyProTyLeuGlyValMetGlyLeuGlyCysAsnPhc 233
Db      670 GGTGGCTCGACACAGAGAAGCTTATCTCTTATACCGGTAAAGATGAAACCTGCAAAATTT 729
QY      234 LysGlnGluAsnValGlyVallyValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
Db      730 TCAGCTGAAAGCTTGGTGTACAGGCTCACTAGTCAACATTAATCTTGGGTGCTGAA 789
QY      254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db      790 GATGAACCTGAAGCATCGGCTTGGATTGTGACGCCAGTAAGCATAGCATTTGAGGTATA 849
QY      274 LysGlyPheAsnLeuTyriLysGlyValTySerSerAspThrCysGlyArgAspPro 293
Db      850 CACTCGTTCGGCTTTTACAGAGTGGAGTTTACCTGATGATGATGATGATGATGATGAT 909
QY      294 MetAspValAsnHisAlaValLeuAlaValGlyTyriGlyValGluAspGlyIleProTy 313
Db      910 ATGATGTGACACACCGGCTTGGCTTGTGTATGAGGTGAGAGCGGTATACCATAT 969
QY      314 TrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyriPheLysMetGlu 333
Db      970 TGGCTTATTAAAGACTCATGGGAGCGGCTGGGCGGCAAAAGGTACTTCAAGATGGAG 1029
QY      334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyriProIleValAla 351
Db      1030 ATGGGGAAGAAATGTGTGTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 1083

RESULT 11
LOCUS   BT000676
DEFINITION Arabidopsis thaliana clone RAFL08-12-G17 (R11172) putative cysteine
SOURCE  proteinase AALP (At5g60360) mRNA, complete cds.
ACCESSION BT000676.1 GI:23397073
VERSION   1
KEYWORDS  Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 1331)
AUTHORS  Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
          Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C.,
          Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
          Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
          Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
          Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
          Ecker,J.R. and Theologis,A.
          Arabidopsis Full Length cDNA Clones
          Unpublished
          2 (bases 1 to 1331)
          Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
          Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
          Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R.,
          Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J.,
          Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
          Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
          Davis,R.W., Ecker,J.R. and Theologis,A.
          Direct Submission
          Submitted (01-OCT-2002) Plant Gene Expression Center, 800 Buchanan
          Street, Albany, CA 94710, USA
          RIKEN Genomic Sciences Center (GSC) members carried out the
          collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
          Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
          Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,

```

```

Hayashizaki,Y. and Shinozaki,K.
The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M.,
Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L.,
Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G.,
Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
Ecker,J.R. and Theologis,A.
Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC)
contributed equally to this work as PIs.
Annotation based on January 2002 version of the Arabidopsis genome
submitted to Genbank.
FEATURES
source      1..1331
             /organism="Arabidopsis thaliana"
             /mol_type="mRNA"
             /db_xref="taxon:3702"
             /chromosome="5"
             /clone="RAFL08-12-G17 (R11172)"
             /ecotype="Columbia"
             /note="This clone is in a modified pBluescript vector
             (PIC-1) as a BamHI/XhoI insert."
gene        1..1331
             /gene="At5g60360"
5'UTR       1..68
             /gene="At5g60360"
misc_difference 2..4
             /gene="At5g60360"
             /note="compared to genomic sequence"
             /replace="gct"
             /protein_id="AA031822.1"
             /db_xref="GI:23397074"
             /translation="MSAKTILSSVVLVLAASAAANIGFDESNPIRMVSDGLREVVEE
             SVSOLIQSRHLSFARFTHRYGKTYQNVEMKLRFSIFKENDLIRSTNKGLSYKL
             GVNQFADLTWQEFQTLGAACNSATLKGSHKTEALPETQWMDGDIYSPVQDQ
             GCSCWTFTTGALAAEAHQAFGKISLSEQLVDCAGAFNNGCNGLPSPQAFYIK
             SNGLDTERAYPTGKDETCFKAENGVQVLNSVNTILGADELHAGVGLVPRVIA
             FEVHSFRLYKSGVTTDSHCGSPMDVNEAVLAVGVEDGVYWLKNSWGADWGDK
             GYFKMENGKMGICATCAYPVVA"
             1146..1331
3'UTR       /gene="At5g60360"
ORIGIN
Alignment Scores:
Pred. No.: 5,85e-107 Length: 1331
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 8 Gaps: 4
US-10-087-714-2 (1-352) x BT000676 (1-1331)
QY 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 69 ATGCTGCGAAAACAATCTCTATCATCATGATGATGATGATGATGATGATGATGATGAT 128
QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 129 GCGGCTAATATCGGATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 198
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 189 GAGGTAGAAGAAATCTGTTCCCAAGATCTTAGGTCAATCTCGTCACGCTTCTCTCTCT 248

```

Qy 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLeuLeuLysAspGlyPhe 77
Db 249 CGTTCCTACCCGATATGTAAAGATATCAAGCTGGAGAGATGAAGCTTCGATTC 308
Qy 78 GlyLeuPheValGluLeuAlaPheAlaArgSerThrAsnArgLysAspLeuSerTyr 97
Db 309 TCGATTTTCAAGGAGATCTTGATTCAGATCCACCAACAAAGAGCTTATCTTAC 368
Qy 98 ThrLeuGlyLeuLeuAlaPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 369 AAATCGGTGTTAATCAATTTGCTGATTCACATGCGCAGAGTTTCAAAGGACCAAGCTT 428
Qy 118 GlyAlaAlaGlnLeuGlnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 429 GGTGCTGCTCAGAACTCTCTGCCACTTTAAAGGCGAGCCACCAAGTCCAGAGAGCT 488
Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGlnGly 157
Db 489 CTTCTCTGAAACAAAGAGCTGGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 548
Qy 158 SerCysGlySer---TTPThrPheSerThrGlyAlaLeuGluAlaLaTyrThrGln 176
Db 549 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
Db 609 GCATTTGGAAGAAAGAAATTTCTCTCTGAGCAACAGCTTGTGGATTTGTGCTGAGCTTTC 668
Qy 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn 213
Db 669 AATAACTGATGTTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 728
Qy 214 GlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCysAsnPhe 233
Db 729 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 788
Qy 234 LysGlnGluLeuValGlyValLysValLysSerLeuLeuLeuLeuGlyValGlu 253
Db 789 TCAGCTGAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 848
Qy 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 849 GATGAACCTGAAGCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908
Qy 274 LysGlyPheLeuLeuLysLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
Db 909 CACTCGTTCGGCTTTTCAAGAGTGGAGTTTACACTGATGATGATGATGATGATGATGATGAT 968
Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyValLeuProTyr 313
Db 969 ATGGATGTGAACCAAGCGGTTTGGCGGTTTGGCGGTTTGGCGGTTTGGCGGTTTGGCGGTT 1028
Qy 314 TrpLeuLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGlu 333
Db 1029 TGGCTTAATAAGAACTCATGGGAGCGGATGGGCGGCAACAAAGGTTACTTCAAGATGGAG 1088
Qy 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProLeuValAla 351
Db 1089 ATGGGGAAGAAACATGTGGTGTGTTGCTACATGCTCATACACCCCGTTGCTGCT 1142
RESULT 12
AY088662
LOCUS AY088662 1352 bp mRNA linear PLN 14-APR-2003
DEFINITION Arabidopsis thaliana clone 8989 mRNA, complete sequence.
ACCESSION AY088662
VERSION AY088662.1 GI:21407436
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1352)
AUTHORS Haas,B.J., Volkovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
TITLE Full-length messenger RNA sequences greatly improve genome
annotation
JOURNAL Genome Biol. 3 (6), RESEARCH0029 (2002)
MEDLINE 22088475
PUBMED 12093376
REFERENCE 2 (bases 1 to 1352)
AUTHORS Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the WS or Lher ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genet carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
FEATURES
Location/Qualifiers
1..1352
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="8989"
80..1156
/codon_start=1
/product="cysteine proteinase AAPL"
/protein_id="AA066984.1"
/db_xref="GI:21617934"
/translation="MSAKTILSSVVLVLAASAAANGIDPESNPIRMVSDGLREVEE
SVSLGSGSRHVSFARTRHYGKXQYQNVEMKLRFSIFKENDLILSTNKKGLSYKL
GVNPDALTWQEFORTKLGAQNCATLKGSHKVEALPDKTOWRSDGIVSPVQDK
GCGSWTFTTGALEAAHYAFGKGISLSEQLVDCAGAFNNGCGLPQSAPEYIK
SNGJLDEKAYPTGKDTCKFSANVGQVLSNVTITGLADELKHAGLVRRPVIK
FEVHSPLKXSGVYTDHSGSTPMDVNVHVLAVGVGVEDGVPYMLIKNSWGDWGDK
GYFKMEMCKNCGIATCASYPVA"
Alignment Scores:
Pred. No.: 5,966-107 Length: 1352
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: Gaps: 4
US-10-087-714-2 (1-352) x AY088662 (1-1352)
Qy 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 80 ATGCTGCGAAGAAACATCTCTATCATCATGTTTGGTGGTTCCTCGCGCATCAGCA 139
Qy 18 ValAlaLeuAlaGlyPheGluGluAspAsnProLleArgSerValThrGlnArgProAsp 37
Db 140 GCGGCTAATATCGGATTCGATGAGTCAAAACCCGATCCGATGCTCCGATGCTCCGG 199

QY 38 SerileGluProAlaIleLeuGlyValleuGlySerCysArgHisAlaPheHisPheAla 57
Db 200 GAGGTAGAGATCTGTTCCAGATCTTAGTCATCTCGTCAGTCTCTCTTCGCT 259
QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLileLysLysArgPhe 77
Db 260 CGCTTCACTCACCGATATGTAATAAGATATCAAGCTGGAGGAGATGAAGCTTCGATTC 319
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
Db 320 TCGATTTTCAAGAGATCTTGAATTCATGATCAGATCCACCAAGAGAGCTTATCTTAC 379
QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTTPGluGluPheArgThrAsnArgLeu 117
Db 380 AAATCGGTGTTAATCAATTTGCTGATTTGATGATGCAAGAGTTTCAAAGGACCAAGCTT 439
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 440 GTGTCTGCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCCCAAGGTCAAGAGCAGCT 499
QY 138 LeuProValThrArgAspTTPArgGluGlnGlnIleValSerProValLysAspGlnGly 157
Db 500 CTTCTGAAACAAAGAGCTGGAGAGAGATGCTATCGTTAGTCCGGTCAAAGATCAGGGA 559
QY 158 SerCysGlySer---TTPThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGln 176
Db 560 GGTGTGGATCTTGTGGACATTCAGCACAACTGGAGCTCTTGAGCGAGCTTACCATCAG 619
QY 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
Db 620 GCATTTGGAAGGAATATCTCTCTGAGCAACAGCTTGTGGATGTCTGGAGCTTTC 679
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn 213
Db 680 AATAACTATGTTGCAATGGTGGCTTCTCTCAAGCCTTTTGAATACATATAAATCCAC 739
QY 214 GlyGlyIleAsnThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhe 233
Db 740 GTGTGGCTCCACACAGAAAGCTTATCTTACCGTAAAGATGAACCTGCAATTT 799
QY 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
Db 800 TCAGCTGAAAGCTTGGTGTACAACTCTCAACTCAGTCAACATTTACTCTGGTGTGAA 859
QY 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 860 GATGAACATGAAGCATCGGTGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 919
QY 274 LysGlyPheAsnLeuTyrIleLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
Db 920 CACTCGTTCGGCTTTTACAAGAGTGGAGTTTACACTGATGATGCTGCTGGAGTACTCCA 979
QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
Db 980 ATGGATGTGAACACCGCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1039
QY 314 TTPLeuIleLysAsnSerTTPGlyThrAsnTTPGlyAspAsnGlyTyrPheLysMetGlu 333
Db -1040 TGCTTTATTAAGACTCATCGGAGCGGATGGGCGGCAAGAGTTATTTCAAGATGAG 1099
QY 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 1100 ATGGGGAAGAACATGTGTGGTATGCTATCATGTGCATCATACCCGGTGTGGCT 1153

RESULT 13
AF083703
LOCUS AF083703 1355 bp mRNA linear PLN 29-OCT-2002
DEFINITION Arabidopsis thaliana clone spe232 unknown mRNA.
ACCESSION AF083703
VERSION AF083703.1 GI:24417303
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 1355)
Stracke,R. and Palme,K.
Signal Peptide Selection derived cDNAs from Arabidopsis thaliana leaves and guard cells
Unpublished
2 (bases 1 to 1355)
Stracke,R. and Palme,K.
Direct Submission
Submitted (12-AUG-1998) Max-Delbrueck-Laboratorium in der Max-Planck-Gesellschaft, Carl-von-Linne-Weg 10, Koeln D-50829, Germany
Location/Qualifiers
1. .1355
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="spe232"
/notes="signal peptide selection derived"
66. .1142
/codon_start=1
/product="unknown"
/protein_id="AA060262.1"
/db_xref="GI:24417304"
/translation="MSAKTILSSVLVILVAASAAANIIGFDESNPIMVSDGLREVEE SVQIIQSRHVLSPFKTHRYGKYQNVEMKLRFSIFKENDLRSTNKKGLSYKL GNQFADLWQEFORTKLGAAQCSATLKGSHKVTAEALPETKDWEDGTVSPVQOQ GCSCSTFTSTGALAAHQAFGKGISLSEQLVDCAGAFNNYCGGLPSPQAFVIK SNGGLTEKAYPTGKDERCKSAENVQVNLNVNITLGADELKHAHGLVPRVSLA FEVHSFLYKSGVYTDHSCGTFPMVDVNHAVLAVGVEDGVPIWLKNSWGDWGDK GYFKMEGKMNCGIATCASYPVVA"
Alignment Scores:
Pred. No.: 5.98e-107 Length: 1355
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 8 Gaps: 4
US-10-087-714-2 (1-352) x AF083703 (1-1355)
QY 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 66 ATGTCTCGGAAAAACAATCTCTATCATCATAGTAGTTTTCGTGGTCTCTCGCCGCGCATCAGCA 125
QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 126 GCSCGCTAATATCGGATTCGATGAGTCAACCCGATCCGATGCTCTCCGATGCTCTCCCG 185
QY 38 SerileGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 186 GAGGTAGAGAATCTGTTCCAGATCTTAGGTCAATCTCGTCACGCTCTCTCTTTCGCT 245
QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLileLysLysArgPhe 77
Db 246 CGCTTCACTCACCGATGATGTAATAAGTATCAGACGTGGAGAGATGAAGCTTCGATTC 305
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
Db 306 TCGATTTTCAAGGAGAATCTTGATTTGATCAGATCCACCAAGAGAGCTTATCTTAC 365
QY 98 ThrleuGlyIleAsnGlnPheAlaAspLeuThrTTPGluGluPheArgThrAsnArgLeu 117
Db 366 AAATCGGTGTTAATCAATTTGCTGATTTGACATGCAAGAGTTTCAAAGGACCAAGCTT 425
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 426 GTGTGCTCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCCACCAAGGTTCACAGAAGCAGCT 485

Qy 138 LeuProValThrArgSerTrpArgGluGlnGlyLeuValSerProValLysAspGlnGly 157
 Db 486 CTTCTCTGAAACAAAGACCTGGAGAGATGGTATCGTTAGTCCGGTCAAGATCAGGGA 545
 Qy 158 SerCysGlySer---TppThrPheSerThrThrGlyValAlaLeuAlaValThrGln 176
 Db 546 GGTTGTGGATCTGCTGGACATTCAGACACAACTGGAGCTCTTGAGGAGCTTACCATCAG 605
 Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
 Db 606 GCATTTTGGAAAAGGAATATCTCTCTGAGCAACAGCTGTGTGATTTGTCTGGAGCTTTC 665
 Qy 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluCysValLysTyrAsn 213
 Db 666 AATAACATATGGTGTCAATGTGGCCCTCTCTCAGCCCTTGAATACATAAATCCAC 725
 Qy 214 GlyGlyLeuAspThrGluGlnThrThrProTyrLeuGlyValMetGlyLeuCysAsnPhe 233
 Db 726 GGTGGCTCGACACAGAGAAAGCTTATCTCTATACCGGTAAAGATGAACCTGCATAATTT 785
 Qy 234 LysGlnGluAsnValGlyValValValValLeuAspSerLeuAsnThrLeuGlyValGlu 253
 Db 786 TCAGCTGAAACAGTTGTGTGACAGTCTCTCACTCAGTCAACATTTACTCTGGGTGCTGAA 845
 Qy 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
 Db 846 GATGAACCTGAAGCATCGGTGGATTGGTACGGCCAGTAAAGCATAGCAATTTGAGGTATA 905
 Qy 274 LysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
 Db 906 CACTCGTTCCGGCTTTTCAAGAGTGGAGTTTACACTGATAGTCACTGTGGAATCTCCA 965
 Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLysProTyr 313
 Db 966 ATGATGTGAACACACCGCGTTTGGCGTTTGGTATGAGTTGAGAGCGGTGACCATAT 1025
 Qy 314 TrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGlu 333
 Db 1026 TGGCTTATTAAAGAACTCATGGGAGCGGATGGGGGACAAAGGTACTTCAAGATGGAG 1085
 Qy 334 LeuGlyLysAsnMetCysGlyValAlaLathCysAlaSerTyrProLysValAla 351
 Db 1086 ATGGGGAGAGACATGTGTGTATTGCTACATGTGCATCATACCCCGTTGGTGGCT 1139
 RESULT 14
 AF360273 1370 bp mRNA linear PLN 18-SEP-2002
 LOCUS Arabidopsis thaliana putative cysteine proteinase AALP (At5g60360)
 DEFINITION mRNA, complete cds.
 ACCESSION AF360273 GI:13430721
 VERSION AF360273.1
 KEYWORDS FULL CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1370)
 Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
 Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
 Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
 Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
 Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Nguyen,M., Palm,C.J., Sakurai,T.,
 Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 Arabidopsis Full Length cDNA Clones
 Unpublished
 2 (bases 1 to 1370)
 Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
 Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
 Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
 Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
 Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Nguyen,M., Palm,C.J., Sakurai,T.,
 Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
 Davis,R.W., Ecker,J.R. and Theologis,A.

Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 Direct Submission
 Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
 Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
 Hayashizaki,Y. and Shinozaki,K.
 The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
 Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
 Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Chen,H.,
 Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
 Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
 this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
 contributed equally to this work as PIs.
 Annotation is based on the January 2002 version of the Arabidopsis
 genome submitted to Genbank.
 Location/Qualifiers
 1..1370
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="RAFL09-18-G19 (RC9257)"
 /ecotype="Columbia"
 /note="This clone is in a modified pBluescript vector
 (FLC-1) as a BamHI/XhoI insert."
 1..1370
 /gene="At5g60360"
 1..39
 /gene="At5g60360"
 /misc_difference 1..2
 /gene="At5g60360"
 /note="compared to genomic sequence"
 /replace="tg"
 /misc_difference 4
 /gene="At5g60360"
 /note="compared to genomic sequence"
 /replace="g"
 40..1116
 /gene="At5g60360"
 /note="contains a papain family cysteine protease domain
 PF00112."
 /codon_start=1
 /evidence=experimental
 /product="putative cysteine proteinase AALP"
 /protein_id="AAK25983.1"
 /db_xref="GI:13430722"
 /translation="MSAKTILSSVVLVAASAAANIGFDESPIRMVSDGLREVSE
 SVSQTIGOSRHVLSFARFTHRYGKYNVEMKLFSGIFKEDLIRSTNKKGLSYKL
 GYNGFADLTWQEFQTKLGAONCSATLKGSHKYTEALPETKDRBDGIVSPVKDQ
 CCGSCWTFSTYTGALAAHYQAFGKGISLSEQLVDGAGANNYCGNGLSPQAFYIK
 SGGSDTETKATPYTGKDETCGSAENVGVQVLSNVITLGADELKXAGLVRRVSLA
 FEVTHSFELYSKYVITDCHGCTPTDVMHVLAVGYGVEDGVFWLTKNSWGDWGDK
 GYFKMEMGKMGATCATCAYPVA"
 1117..1370
 /gene="At5g60360"
 1354
 /misc_feature
 /gene="At5g60360"
 /note="artifact within poly A tail"
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,068-107 Length: 1370
 Score: 1279.50 Matches: 240

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 22:16:09 ; Search time 369 Seconds

(without alignments)
4052.482 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAAKLFLFLVLSALVAL.....ELGKMGVATCASYPIVAV 352

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10087714/runat 26032004 113452 923/app query.fasta_1.519
-DB=N Geneseq 29Jan04 -QMT=fascap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087714.ccgcn 1 1 352 @runat 26032004 113452 923 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04: *
1: geneseq1980s: *
2: geneseq1990s: *
3: geneseq2000s: *
4: geneseq2001as: *
5: geneseq2001bs: *
6: geneseq2002as: *
7: geneseq2003as: *
8: geneseq2003bs: *
9: geneseq2003cs: *
10: geneseq2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1819	97.5	1071	ADC61376	ADC61376 DNA encod
2	1310.5	70.3	1442	AAT35111	Aat35111 Enhanced
3	1300.5	69.7	1407	AAP44802	Aaf44802 Cysteine
4	1300.5	69.7	1407	ADB94743	ADB94743 Programme
5	1279.5	68.6	1358	AAC37877	Aac37877 Arabidops
6	1279.5	68.6	1370	AAAC48945	Aac48945 Arabidops
7	1270	68.1	1467	AAV82457	AAV82457 Triticum
8	1265.5	67.9	1077	AAAC51435	AAAC51435 Arabidops

C	9	1263	67.7	1282	6	ABN98252	Arabidops
	10	1259.5	67.5	1363	3	AAC40628	Arabidops
	11	1227	65.8	1388	2	AAT35100	Up-regula
	12	1148	61.6	1652	9	ADB94877	Programme
	13	1108	59.4	2149	3	AAC44178	Arabidops
	14	1032.5	55.4	1072	3	AAC44535	Zea mays
C	15	846	45.4	618	8	ACL19838	DNA clone
	16	845.5	45.3	888	9	ADB94874	Programme
C	17	844.5	45.3	675	8	ACL19836	DNA clone
C	18	841	45.1	662	8	ACL19837	DNA clone
	19	796	42.7	1362	7	ABT41873	Toxicity
	20	796	42.7	1362	9	ADB52643	Primary r
C	21	793	42.0	590	8	ACL19847	DNA clone
C	22	774	41.5	565	8	ACL19840	DNA clone
C	23	765.5	41.0	644	8	ACL19841	DNA clone
C	24	763	40.9	1399	6	ABR84545	Human cDN
	25	759	40.7	1433	9	ADD18512	Human pro
C	26	759	40.7	1475	4	AAI93479	Human pol
C	27	759	40.7	1475	4	AAI93494	Human pol
	28	759	40.7	1490	3	AAF18016	Lung canc
	29	759	40.7	1503	6	ABZ35301	Human gen
	30	759	40.7	1503	9	ADE53672	Human pro
C	31	748.5	40.1	630	8	ACL19848	DNA clone
C	32	736.5	39.5	595	8	ACL19839	DNA clone
	33	735	39.4	588	7	ABX56737	Arabidops
C	34	723	38.8	536	8	ACL19835	DNA clone
C	35	711.5	38.2	604	8	ACL19843	DNA clone
	36	710	38.1	1109	5	AA570355	DNA encod
	37	691	37.1	1106	6	ABZ35041	Human gen
	38	691	37.1	1106	6	ADE84838	Farnesyl
C	39	676.5	36.3	550	8	ACL19846	DNA clone
C	40	676.5	36.3	578	8	ACL19845	DNA clone
C	41	667.5	35.8	545	8	ACL19849	DNA clone
C	42	658	35.3	714	8	ACL19871	DNA clone
	43	656.5	35.2	677	4	AAF44806	Cysteine
	44	656.5	35.2	677	9	ADB94747	Programme
	45	656.5	35.2	1075	2	AAQ70268	Fasciola

ALIGNMENTS

RESULT 1

ADC61376

ID ADC61376 standard; DNA; 1071 BP.

XX AC ADC61376;

XX DT 18-DEC-2003 (first entry)

XX DE DNA encoding Vanilla planifolia 4-hydroxybenzaldehyde synthase (4HBS).

XX DE Vanillin production; plant; vanillin biosynthesis;

XX KW 4-hydroxybenzaldehyde synthase; 4HBS; gene; ds.

XX OS Vanilla planifolia.

XX FH Key Location/Qualifiers

XX CDS 1..1071

XX FT /*tag= a

XX FT /product= "4HBS"

XX FT /transl_except= (pos:481..489,aa:Ser-Tip)

XX FT /transl_except= (pos:538..546,aa:Gly-Ser)

XX FT /transl_except= (pos:547..555,aa:Thr-Leu)

XX FT /transl_except= (pos:604..612,aa:Cys-Gly)

XX FT /transl_except= (pos:694..696,aa:Met)

XX XX US2003070188-A1.

XX PN 10-APR-2003.

XX PD 28-FEB-2002; 2002US-00087714.

XX PF 15-JUL-1997; 97US-0052604P.

PI Farrell A, John I;
 XX WPI; 1995-131361/17.
 XX
 XX DNA constructs which modify expression of senescence-related genes -
 XX useful to accelerate or inhibit senescence in crop plants.
 XX
 XX Claim 16; Page 55-56; 78pp; English.
 XX
 CC The sequences given in AAT35095-133 are senescence related genes and
 CC fragments which were used in the construct of the invention which
 CC modifies the expression of at least one senescence related gene. Using
 CC these constructs senescence may be inhibited or accelerated in plants
 CC including tomato, lettuce, cabbage, banana, strawberry, wheat, maize,
 CC rice, rape or soybean. Delayed senescence may indirectly prolong the life
 CC of the plant, increase yield, increase protein content of fruits, improve
 CC quality of leafy vegetables, improve tolerance to disease, drought or
 CC other stresses. Increased senescence may more rapidly break down unwanted
 CC plant material and so avoid the use of desiccants on crops. This sequence
 CC represents the senescence-enhanced clone, SEE1, which is also known as
 CC clone p16.4. SEE1 is a cDNA of approx. 1.7 kb encoding a mRNA of approx.
 CC 1.2 kb. The mRNA encoded by SEE1 increases in abundance during maize leaf
 CC senescence. SEE1 shows homology to genes for two thiol proteases, oryzain
 CC gamma from rice and aleurain from barley. The clone SEE1 may therefore
 CC encode a protease. This is supported by the finding that a 1 kb DNA
 CC fragment is amplified by PCR when one of the primers used is derived from
 CC a region which is conserved in a range of thiol proteases. The predicted
 CC amino acid sequence of the protein encoded by SEE1 suggests that the
 CC protein is 360 amino acids long with a mol. wt. of 39 kD. There are
 CC conserved motifs in the amino acid sequence, a putative vacuolar signal,
 CC a Cys active site, a His active site and an Asn active site. The
 CC predominantly hydrophilic protein sequence suggests that the protein is
 CC soluble. SEE1 is deposited as NCIMB 40582
 XX
 SQ Sequence 1442 BP; 314 A; 408 C; 393 G; 327 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8,418-137 Length: 1442
 Score: 1310.50 Matches: 257
 Percent Similarity: 78.61% Conservative: 26
 Best Local Similarity: 71.39% Mismatches: 68
 Query Match: 70.27% Indels: 9
 DB: 2 Gaps: 6
 US-10-087-714-2 (1-352) x AAT35111 (1-1442)
 QY 1 MetAlaAlaLysLeuLeuPhePheLeuValSerAlaLeuSerValAlaLeu 20
 DB ATGGGCCCCACGGCGCTGCTCTCTCGCCGTCGTCGCCCTCGCGCCACCGCGCGCG 137
 QY 21 Ala-----GlyPheGluAspAsnProIleArgSerValThrGlnArgProAspSer 38
 DB GCGCAACTCGCGCTCGCGGACTCCAAACCGGATCCGCCCGCTCACCGACCGCGCGCTCC 197
 QY 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
 DB GCGCTCGAGTCCACGGCTTCGCCCGCTCGCGCCACCGCGCGCGCTCGCTTCGCA 257
 QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPhe 77
 DB CGCTTCGCGCTCAGGTACGGCAAGAGCTACGAGAGCGCGCGCGGAGTCCACAGCGGTTT 317
 QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspIleSerTyr 97
 DB AGGATCTTCTCCGAGAGCTCCAGCTGTCGCTCCACCAACCGCAAGGCTCTCTCTAC 377
 QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
 DB CGCTTCGCGATCAACCGCTTCGCGACATGAGCTGGAGGAGTTCGCTGCGACCGCGCTC 437
 QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
 DB GGTGAGGCCAGAACTGCTCGGCCAGCTTACCGGCAACCAACCGGATCGCGCGCGCC 497

QY 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAsp 155
 DB GTTTCGCGTTCGCGAGAGCAAGAGCTGGAGGAGGATGGGATTTGTGAGCCCGAGTAAAC 557
 QY 156 GlnGlySerCysGlySer---TyrThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174
 DB CAGGCGCACTGTGATCATCTGTCGACTTCTGAGCACTTCTGTCGCTTGGGGTGTATAT 617
 QY 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
 DB ACCCAGCAACTGGCAAGCCCATCTCTCTCTGAGCAACAGCTTGTGACTGTGGTTT 677
 QY 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
 DB GCATTCAACATTTTCGATGCAACGAGGCGCTTCATCCGAGGCCCTTTGATACATCAAA 737
 QY 212 TyrAsnGlyGlyIleAspThrGluInThrTyrProTyrLeuGlyValMetGlyIleCys 231
 DB TACAATGGTGGCTTGCACCTGAGGAATCTTACCCCTTACCAGGTGTCAATGGAATCTCC 797
 QY 232 AsnPheLysGlnGluAsnValGlyValLysValLysSerIleAsnIleThrLeuGly 251
 DB AAGTTTAAAGAAATGAGAAATGTTGAGTCAAGGTTTTCGACTCGGTTAACATCACCCCTGGT 857
 QY 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 DB GCTGAGGATGAATGAAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 917
 QY 272 ValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArg 291
 DB GTGATCACTGTTTCAGGCTGTACAGAGCGGAGTTTACACTAGCAGCCACTTGTGAACT 977
 QY 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311
 DB ACACCGATGGATGTGAACACGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1037
 QY 312 ProTyrTrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
 DB CCTACTGCTCATCAAGAACTCATGGGCGCTGACTGGGGGTGATGAGGGTTTACTTCAAG 1097
 QY 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
 DB ATGGAATGGGCAAGAACTGTCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1157
 RESULT 3
 AAF44802
 ID AAF44802 standard; cDNA; 1407 BP.
 XX
 AC AAF44802;
 XX 27-MAR-2001 (first entry)
 DT
 XX Cysteine protease coding sequence #16.
 DE Cell death modulator; programmed cell death; PCD; apoptosis;
 KW forestry plant; ss.
 XX Eucalyptus grandis.
 OS
 XX WO200075331-A1.
 FN
 XX 14-DEC-2000.
 ED
 XX 02-JUN-2000; 2000WO-NZ000086.
 PF
 XX 04-JUN-1999; 99US-00325932.
 PR
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX Flinn B, Lasham A;
 PI
 XX

CC reverse complements, reverse sequences, or sequences having 75, 90 or 95
 CC % sequence identity to or that hybridise under stringent hybridisation
 CC conditions to one of the 145 sequences. The methods and compositions of
 CC the present invention to do with (1) are useful for modulating programmed
 CC cell death and thereby altering the development cycle of plant cells, and
 CC altering plant development. This sequence encodes a protein associated
 CC with the programmed cell death pathway.

XX SQ Sequence 1407 BP; 349 A; 345 C; 377 G; 336 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,08e-135 Length: 1407
 Score: 1300.50 Matches: 247
 Percent Similarity: 79.77% Conservative: 33
 Best Local Similarity: 70.37% Mismatches: 60
 Query Match: 69.73% Indels: 11
 DB: 9 Gaps: 5

US-10-087-714-2 (1-352) x ADB94743 (1-1407)

Qy 9 LeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAspAsnPro 28
 Db 91 CTCCTCTCGTCGCGTCGTCGTCGCGCGCGTCGAGCTTCGAGAGTCCAAACCC 150
 Qy 29 IleArgSerValThrGlnArgProAsp-----SerIleGluProAlaIleLeu 44
 Db 151 ATCCGCTCTTC-----CCGACGGGGCTCCGGACCTCGAGCTCCATCGTC 201
 Qy 45 GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgTyrGly 64
 Db 202 CAGATCGTCGCGCCACCCCGCCAGCTTCTCTCGCCGCTTCGCCAACAGGTATGG 261
 Qy 65 LysSerTyrGlySerGluGluGluIleLysArgPheGlyIlePheValGluAsnLeu 84
 Db 262 AAGAGTACGAGACCGCGGAGGAGATCAAGTCGGTTCGAGATCTTCAGGGAGATCTC 321
 Qy 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104
 Db 322 AAGTTGATCCGATCCACCAACAGAGGGCTTGCCCTACACCTCGTGTCAATAGTT 381
 Qy 105 AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
 Db 382 GCTGATTGGAGCTGGGAGGAGTTCAGGACGACAGATCGGAGCTGCTCAAACTGCTCT 441
 Qy 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
 Db 442 GCCACCAACGAGGACACCAAGCTCACCAGAGAGCTTTCGCGAGTGAAGACTGG 501
 Qy 145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer----TrpThr 163
 Db 502 AGAGAAAAGGGCATTTGAAGCCCAATTAAGATCAGGGGCACCTGTGATCTTCTGACT 561
 Qy 164 PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
 Db 562 TTCAGTACCACTGGAGCTTTGAGGCTGCTTATCACCAGCATTCGGAAACAAATCTCT 621
 Qy 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
 Db 622 CTGTCTCAGCAGCAGCTGCTGAGCTGTGCTGGGCTTCAACAACTTGGATGTAGTGT 681
 Qy 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
 Db 682 GGACTGTCATCCCAAGGCTTTGAGTAGTCAAGTACCAAGTGGGCTTGTATACCGAGAA 741
 Qy 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
 Db 742 GCATATCTTATACCGCAGTGGATGGTAGTGCATAATCTTCGGCTGATATGTTGTGTC 801
 Qy 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
 Db 802 CAAGTGTCTGACTCTGTAACATCACCTTGGGCTGCTGAGGATGAACCTAAAGCATGCA 861
 Qy 261 GlyLeuValArgProValSerValAlaPheGluValLysGlyPheAsnLeuTyrLys 280

Db 862 GCCTTCGTCGCGCAGTCAGTGTGGCAATTCAGGTCGTAAGACTTCAGATTGTACAAG 921
 Qy 281 LysGlyValTyrSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
 Db 922 TCGGTGTCTACACAGGAGTACATCGCGTAGCATCTTCATGGATGGAACCATCTGTT 981
 Qy 301 LeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTyrLeuIleLysAsnSerTyr 320
 Db 982 CTCGCTGTGTGTTATGAGATTGAAGATGGTTCGTTTCGTTCTCATCAAGAAATCTCTGG 1041
 Qy 321 GlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGly 340
 Db 1042 GGAGCAGATCGGGTGACCGAGATCTCAAGATGGAGATGGGAAGAACATGTGTGGA 1101
 Qy 341 ValAlaThrCysAlaSerTyrProIleValAla 351
 Db 1102 GTGCTACTTGTGCATCATACCTCTGTGTGGCC 1134

RESULT 5
 AAC37877
 ID AAC37877 standard; DNA; 1358 BP.
 XX AAC37877;
 AC AAC37877;
 DT 17-OCT-2000 (first entry)
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 18989.
 DE Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-00301439.
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144332P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 23-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-014926P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160816P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.

PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Alignment Scores:
Pred. No.: 2,33e-133 Length: 1358
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 3 Gaps: 4

US-10-087-714-2 (1-352) x AAC37877 (1-1358)

QY 1 MetAlaAlaLysLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
DB 81 ATGCTCGGAAACAACTATCATCATAGTTTGGTGGTTCCTCGTCGCCGATCAGCA 140
QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
DB 141 GCGGCTAATATCGGATTCGATGAGTCAAAACCGATCCGAATGGTCTCCGATGGTCTCCGG 200
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
DB 201 GAGGTAGAAAGATCTGTTCCAGATCTTAGTCAATCTCTCAGTCTCTCTCTTCGCT 260
QY 58 ArgPheAlaArgArgTyGlySerTyGlySerGluGluGluLeuLysLysArgPhe 77
DB 261 CCGTTCACCTACCGATATGTAAGATATCAGAACGTCGAGAGATGAAGCTTCGATTC 320
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTy 97
DB 321 TCGATTTTCAAGGAGAACTTGAATTTGATTCAGATCCCAACAGAAAGGCTTATCTTAC 380
QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 381 AAACCTCGGTGTTAATCAATTTGCTGATTTGATGTCGCAAGAGTTTCAAGACCAAGCTT 440
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
DB 441 GGTGTCTGTCAAGATCTGTCTGCCACTTAAAGGCGCAGCCCAAGGTCACAGAAGCAGCT 500
QY 138 LeuProValThrArgAspTrpArgGlnGlnGlyIleValSerProValLysAspGlnGly 157
DB 501 CTTCTCTGAAACAAAGACTGGAGAGAGATGGTATCGTTAGTCCGGTCAAAGATCAGGA 560
QY 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyThrGln 176
DB 561 GGTGTGGATCTTGTGGACATTCAGCAACACTGGAGCTCTTGAGGCGAGCTTACCATCAG 620
QY 177 LeuThrGlySer-----ThrIleSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
DB 621 GCATTTGGAAAGGAATATCTCTCTGAGCAACAGCTTGGATTTGGTGGAGCTTTC 680
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyThrValLysTyAsn 213
DB 681 AATAACTATGTTGCAATGGTGGCCTTCTCTCAAGCCTTTGAATACATAAAATCCAAC 740
QY 214 GlyGlyIleAspThrGluGlnThrTyProTyLeuGlyValMetGlyIleCysAsnPhe 233
DB 741 GGTGGCTCGACAGAGAAAGCTTATCTTATACCGGTAAGATGAACCTGCAATTT 800
QY 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyValGlu 253
DB 801 TCAGCTGAAACGTTGGTGTACAAGTCTCAACTCAGTCAACATTTACTTCTGGTGTGAA 860
QY 254 AspGluLeuLysPheHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
DB 861 GATGAACGTGAACATGCGGTTGGATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 920
QY 274 LysGlyPheAsnLeuTyLysLysGlyValTySerSerAspThrCysGlyArgAspPro 293
DB 921 CACTCGTTCGGCTTTACAGAGTGGAGTTTACACTGTATGATCTACTGTGGAGTACTCCA 980

QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyThrGlyValGluAspGlyIleProTy 313
DB 981 ATGGAATGTAACACCGCGTTTGGCGTTTGGTATGGAGTTGAGACCGTGTACCATAT 1040
QY 314 TrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyThrPheLysMetGlu 333
DB 1041 TGGCTTATTAAAGAACTCATGGGAGCGGATTCGGGCGACAAAGGTTACTTCAAGATGGAG 1100
QY 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyProIleValAla 351
DB 1101 ATGGGAAGAACAATGTGTGTATTGTCTATCATGTGCATCATACCCCGTTGTGGCT 1154

RESULT 6
AAC48945
ID AAC48945 standard; DNA; 1370 BP.
XX
AC AAC48945;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59352.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 23-MAR-1999; 99US-0126264P.
PR 23-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 23-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140921P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.

PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149388P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 13-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160985P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 2,36e-133
Score: 1279.50
Percent Similarity: 77.93%

Length: 1370
Matches: 240
Conservative: 39

Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 3 Gaps: 4

US-10-087-714-2 (1-352) x AAC48945 (1-1370)

```
QY 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
DB 80 ATGTCGCGAAACAACTCATCATCATGTTTGGTGGTTCGTCGCGCATCAGCA 139
QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
DB 140 GCGGCTAATATCGGATTCGATGAGTCAAAACCGATCGAATGGTCTCCGATGGTCTCCGG 199
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
DB 200 GAGGTAGAGAATCTGTTCCAGATCTTAGTCAATCTCGTCACGTTCTCTCTTTCGCT 259
QY 58 ArgPheAlaArgArgTyrGlySerGluGluGluIleLysLysArgPhe 77
DB 260 CGCTTCACTACCGATATGTTAAAGATATCAGACGTTGGAGGAGATGAAGCTTCGATTC 319
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
DB 320 TCGATTTTCAGGAGAACTCTGATTTGATCAGATCCACCAACAAAGAAAGGCTTATCTTAC 379
QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 380 AAATCGGTGTTAATCAATTTGCTGATTTGATGCGCAAGAGTTTCAAGACCAAGCTT 439
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
DB 440 GGTGCTGCTCAGAACTGCTCGCACTTTAAAGGCGCAGCCACAAAGGTCACAGAAGCAGCT 499
QY 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGly 157
DB 500 CTTCCTGAAACAAAGATCGAGAGAGATGATATCGTTAGTCCGGTCAAGATCAGGGA 559
QY 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGln 176
DB 560 GGTGCTGATCTTGTGTCGACATTCAGCAACTGGAGCTCTTGAGGCAAGCTTACATCAG 619
QY 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
DB 620 GCATTTGGAAGAAGAAATTTCTCTCTCAGCAACAGCTTGTGGATTTGCTGGAGCTTTC 679
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn 213
DB 680 AATACTATGTTGCAATGGTGGCTTCTCTCAAGCTTTGAATACATCAATCCAAAC 739
QY 214 GlyGlyIleAspThrGluGlnThrTrpProTyrLeuGlyValMetGlyIleCysAsnPhe 233
DB 740 GGTGGCTCGACACAGAGAAAGCTTATCTTATACCGGTAAAGATGAACCTGCAAAATTT 799
QY 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
DB 800 TCAGCTGAAACGTTGGTGTACAGTCTCACTCACTCAACATCTACTCTGGGTGCTGAA 859
QY 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
DB 860 GATGAATGAAGCATCGGTTGGATTTGATCGGCCAGTAAAGCATAGCATTTGAGGTTATA 919
QY 274 LysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
DB 920 CACTCGTTCGCGCTTTTACAAGAGTGGAGTTTACACTGATAGTCTACTGTGGAGTACTCCA 979
QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
DB 980 ATGGATGTGAACACACCGGTTTGGCCGTTGGTTATGGAGTTGAAGACGGTGTACCATAT 1039
QY 314 TrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGlu 333
DB 1040 TGGCTTATTAAAGACTCATGGGAGCGGATGGGGCGCAAAAGGTTACTTCAAGATGGAG 1099
```

```
QY 334 LeuGlyIysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
DB 1100 ATGGGGAAGAACATGTGTGTATTGTATCATGTGTCATACCCCGTGTGGCT 1153
```

RESULT 7

AAV82457
ID AAV82457 standard; DNA; 1467 BP.

AC AAV82457;

DT 16-MAR-1999 (first entry)

DE Triticum sp. cysteine proteinase #2 encoding DNA.

KW Triticum; wheat seed; cysteine proteinase; gluten; baking; ds.

OS Triticum sp.

PH Key Location/Qualifiers
CDS 47..1144
FT /*tag= a

XX JP10327886-A.

XX PN 15-DEC-1998.

XX PD 27-MAR-1998; 98JP-00098140.

XX PF 31-MAR-1997; 97JP-00114946.

XX PR (SHOS) SHOWA SANGYO CO.

XX PS WPI; 1999-109255/10.

XX PP P-PSDB; AAW89557.

XX PT New DNA coding cysteine proteinase originating from wheat seed - useful

XX PT for improving gluten for use in bakery process.

XX PS Claim 2; Page 16-18; 29pp; Japanese.

XX CC The present sequence encodes a cysteine proteinase isolated from wheat

XX CC seed (Triticum sp.). The cysteine proteinase is useful for improving

XX CC gluten for use in the bakery process

XX SQ Sequence 1467 BP; 348 A; 423 C; 389 G; 305 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 3 04e-132 Length: 1467
Score: 1270.00 Matches: 246
Percent Similarity: 78.12% Conservative: 36
Best Local Similarity: 68.14% Mismatches: 67
Query Match: 68.10% Indels: 12
DB: 2 Gaps: 6

US-10-087-714-2 (1-352) x AAV82457 (1-1467)

QY 3 AlaLysLeuLeuPhePheLeuLeuValSerAlaLeuSerValAlaLeuAla--- 21

DB 56 GCGCGATCTCTCTCTCGCGCTCGCGCCGCCACCGCCACCGCTCGCGCTCC 115

QY 22 -----GlyPheGluGluAspAsnProIleArgSerValThrGlnArgPro 36

DB 116 TCCTCTCTCTCTCTCTCTCTCTCAACCCGATCCGCGCCGCTCCGCGCGCGC 175

QY 37 AspSer---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHis 55

DB 176 GCTTCACGCTCGAGTCCCGGCTCTCGCGCGCTCGCGCGCGCACCGCCCTCCGC 235

QY 56 PheAlaArgPheAlaArgTyrGlyLysSerTyrGlySerGluGluGluLysLys 75

DB 236 TTCGCGGCTTCGCGCTCAGGTACGACAGAGTACAGAGCGCGCGAGTCCGATG 295

QY 76 ArgPheGlyLeuPheValGluAsnLeuAlaPheIleArgSerThrAsnArgIysAspLeu 95
DB 296 CGGTTCCGGATCTTCGAGAGCCTCGAGGAGTCCGCTCCACCAACCGGAAGGCGCTA 355
QY 96 SerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsn 115
DB 356 TCTACCGCCTCGGATCAACCGCTTCGGACATGAGTGGAGAGTTCAGGCGACC 415
QY 116 ArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAsp 135
DB 416 CGGCTCGGCGCGCGAGACCTCGCTCGGCACGCTCGCGGCAACCACTGATGCGGCAC 475
QY 136 --GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValIys 154
DB 476 GCGCGCGCCTCCGGAGACCAAGACCTGGAGGAGGATGGATCGTTAGCCCGTAAA 535
QY 155 AspGlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAla 173
DB 536 GACCACTCCCATTCGGCTCTCGGATCGAGCTTACGACCTACTGCTGCACTTTGAGCACA 595
QY 174 TyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAla 191
DB 596 TATACTCAGGCACCTCGAAGAACATCTCTCTTTCTGACCAACACTGTTGACTGTGCC 655
QY 192 SerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVal 210
DB 656 GCGGATTCATTAATTCGATCGAGCGAGGCTTCATCCAGGCATTTGAGTACATC 715
QY 211 LysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIle 230
DB 716 AAATACATGGAGGATCGATACCGAGGAGTCTACCTTACAGGGTGTCTATGGCGTC 775
QY 231 CysAsnPheLysGlnGluAsnValGlyValIysValIleAspSerIleAsnIleThrLeu 250
DB 776 TGCACTTACAGCTCAAAATGCGGTAGTTGAGTTTGGCTCTGTTAAACATCACACTG 835
QY 251 GlyAlaGluAspGluLeuIysHisAlaValGlyLeuValArgProValSerValAlaPhe 270
DB 836 ATGCTGAGAGCAACTGAAACCGCGCTGGTGGTTCGCCAGTGGATGTTGCCCTT 895
QY 271 GluValValIysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGly 290
DB 896 GAGTGATCAACGGTTTCAGGCAGTACAGAGCGGAGTTTACATAGCGACCATTCGCGT 955
QY 291 ArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGly 310
DB 956 ACTACACCTGATGACGTGAACCAACCGCGCTTCGGCGGTGCGGTACGGGTGGAGAACGCG 1015
QY 311 IleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyValAspAsnGlyTyrPhe 330
DB 1016 GTCCCGATTGGCTCATCAAGAACTCTGGGGCGCGACTGGGTGACAACTGTTACTTC 1075
QY 331 LysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleVal 350
DB 1076 AAGATGGAATGGCAAGAACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1135
QY 351 Ala 351
DB 1136 GCG 1138

RESULT 8
AAC51435
ID AAC51435 standard; DNA; 1077 BP.
XX
XX AAC51435;
AC
XX 18-OCT-2000 (first entry)
DT
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 68510.
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 68510.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.

XX OS
XX Arabidopsis thaliana.
FN EPI033405-A2.
XX PD
XX 06-SEP-2000.
XX PF
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.

QY 98 ThrLeuGlyIleAsnGlnPheAlaSerLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 301 AAACCTCTCTCTTAATCAGTTTGTGACTTGTGACATGCGAAGATTTTCAAGATACAGCTT 360
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
DB 361 GAGAGTCTCTCAAACTGTTCTGCTACTTTAAAGGGTAGCCACAAGTCACTGAAGCTACA 420
QY 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValIleAspGlnGly 157
DB 421 GTTCCAGACACAAAGATTGGAGAGAAGATGTTTGTAGCCCTGTGAAAGAACAGGGA 480
QY 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaThrGln 176
DB 481 CATTGTGGATCTTGTGGACATTTAGCACACTGGAGCTCTTGAAGAGCTTACCATCAA 540
QY 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
DB 541 GCATTTGGAAGGAATATCTTTGTCGAGCAACAGCTTGTGGATTGTGCTGCTACTTC 600
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluThrValIleSerThr 213
DB 601 AATAACTTTGGTTGTCTGTTGACCTTCTTCAAGCCCTTTGAATACATTAATAATACAC 660
QY 214 GlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnZhe 233
DB 661 GTTGGGCTCGACACGAGGAGGCTTATCTTACACCGGAAAGACGGTGGCTGCAATTT 720
QY 234 LysGlnGluAsnValGlyValIleValIleAspSerIleAsnIleThrLeuGlyIleGlu 253
DB 721 TCACGCAAAACATCGGTGTACAAAGTCCGTGACTCTGTCAACATTACCTGGTGGTCAGAA 780
QY 254 AspGluLeuIleHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
DB 781 GATGAAGTGAAGACGCGGTGGTGGTGAAGCCAGTGGTGGTGGTGGTGGTGGTGGT 840
QY 274 LysGlyPheAsnLeuTyrIleValIleValIleValIleValIleValIleValIleVal 293
DB 841 CATGAATTCAGGTTTATAGAGGAGGAGTGTCTAGCAATACATGCTGGTGAACACTCA 900
QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
DB 901 ATGATGTAAACCATCGAGTGTAGCAGTGTGTATGGAGTTGAAGACGACGCTCCGATAC 960
QY 314 TrpLeuIleIleAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheIleValMetGlu 333
DB 961 TGGCTTATAAAGAACTCATGGGAGGTGAATGGGAGACAATGGCTACTTCAAGATGGAA 1020
QY 334 LeuGlyIleAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
DB 1021 ATGGGAAGAACATGT 1074
RESULT 9
ID ABN98252/c
XX ID ABN98252 standard; DNA; 1282 BP.
AC ABN98252;
XX
DT 01-AUG-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 20.
XX
KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
KW disease; crop; thale cress; tolerance factor; insect; pathogen;
KW nutrition; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002023281-A1.
XX
XX 21-FEB-2002.
XX

26-JAN-2001; 2001US-00770445.
27-JAN-2000; 2000US-0178472P.
(GORL/) GORLACH J.
(ANY/) AN Y.
(HAMI/) HAMILTON C M.
(PRIC/) PRICE J L.
(RAIN/) RAINES T M.
(YUY/) YU Y.
(RAME/) RAMEAKA J G.
(PAGE/) PAGE A.
(MATH/) MATHW A V.
(LEDF/) LEDFORD B L.
(WOES/) WOESSNER J P.
(HAAS/) HAAS W D.
(GARC/) GARCIA C A.
(KRICK/) KRICKER M.
(SLAT/) SLATER T.
(DAVI/) DAVIS K R.
(ALLE/) ALLEN K.
(HOFF/) HOFFMAN N.
(HURB/) HURBAN P.
Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
Hurban P;
WPI: 2002-403163/43.
New Arabidopsis thaliana nucleic acid for identifying homologous genes,
producing compositions that modulate the expression or function of its
encoded protein, and mapping functional regions of protein.
Claim 1; SEQ ID NO 20; 49pp + Sequence Listing; English.
The invention relates to an Arabidopsis thaliana nucleic acid (I)
comprising a sequence capable of hybridising under stringent conditions
to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
given in the specification or its fragment. A polypeptide (II) encoded by
(I), a transgenic plant (III) comprising an exogenous nucleic acid or a
genetically modified cell (IV) comprising an exogenous nucleic acid, is
useful for screening a candidate agent for its biological effect. (I) is
useful in identifying homologous or related genes, in producing
compositions that modulate the expression or function of its encoded
protein, mapping functional regions of the protein and in studying
associated physiological pathways. (I) is also useful for the genetic
manipulation of cells, particularly plant cells. (I) is also useful in
screening assays of various plant strains to determine the strains that
are best capable of withstanding a particular disease or environmental
stress. (II) and (III) are useful for screening of biologically active
agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
pathways. The screened agents are useful in improved methods of treating
crops to prevent or treat disease. (II) are also useful in screening
programs to identify agents that mimic or enhance the action of tolerance
factors. Such agents are useful in improved methods of treating crops to
enhance their tolerance to environmental stress. (I) is also useful for
enhancing or inhibiting production of a biosynthetic product in a plant.
(III) is useful for identifying other mediators that may induce
expression of proteins of interest, for establishing the extent to which
any specific insect and/or pathogen is responsible for damage to a
particular plant, for identifying other mediators that enhance or induce
tolerance to environmental stress, for identifying factors involved in
biosynthetic pathways of nutritional, commercial or medicinal value
for identifying productions of nutritional, commercial or medicinal
value. (IV) is useful in the study of genetic function and regulation,
for alteration of the cellular metabolism and for screening compounds
that may affect the biological function of the gene or gene products.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from USPTO
at seqdata.uspto.gov/sequence.html?DocID=999909770445

SQL Sequence 1282 BP; 367 A; 314 C; 234 G; 367 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.53e-131 Length: 1282
 Score: 1263.00 Matches: 235
 Percent Similarity: 78.39% Conservative: 37
 Best Local Similarity: 67.72% Mismatches: 71
 Query Match: 67.72% Indels: 4
 DB: 6 Gaps: 3

US-10-087-714-2 (1-352) x ABN98252 (1-1282)

QY 9 LeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAspPro 28
 DB 1252 CTATTGATTCCTTTGGCGGCGGAGTAAGAGATCGGATTCGATGATCAATCCA 1193
 QY 29 IleArgSerValThrGlnArgProAspSerIleGluProAlaIleLeuGlyValIleuGly 48
 DB 1192 ATCAAAATGGTCTCCGCAATCTCCATGAGCTTGAAGACACCGTCTGCATGCTTAGGA 1133
 QY 49 SerCysArgHisAlaPheHisPheAlaArgPheAlaArgAIGTyrglySerTyrgly 68
 DB 1132 CAGTCCCGTCATGTTCTATCTTTCTGCTCACTCAGGTATGGGAAGAATATCAG 1073
 QY 69 SerGluGluGluIleLeuLysArgPheGlyIlePheValGluAsnLeuAlaPheIleArg 88
 DB 1072 AGTGTGGAGGAGATCAAGTAAAGATCTCTGTTTAAAGGAGAACTTTGATTTAATCAGA 1013
 QY 89 SerThrAsnArgLysAspLeuSerTyrrThrLeuGlyIleAsnGlnPheAlaAspLeuThr 108
 DB 1012 TCACATAAAGAAAGGCTTATCTATAAATCTCTCTTAAATCACTTTGCTGACTTGACA 953
 QY 109 TrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSerAlaAlaThrAlaHis 128
 DB 952 TGGCAAGAGTTTCAAGATACAAAGCTGGAGCTGTCAAACTGTCTGCTACTTTAAAG 893
 QY 129 GlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTrpArgGluGlnGly 148
 DB 892 CGTAGCCAGAGATCACTCAAGCTACAGTTCAGACACAAAAGATTGGAGAGAAGTGGT 833
 QY 149 IleValSerProValLysAspGlnGlySerCysGlySer---TrrThrPheSerThrThr 167
 DB 832 ATGTGATGCCCTGTGAAGACAGGACATTTGGATCTTTGGACATTTAGCACAACT 773
 QY 168 GlyAlaLeuGluAlaAlaTyrrThrGlnLeuThrGlySer-----ThrLeuSerGluGln 185
 DB 772 GGAGCTCTTGAAGCAGCTTACCATCAAGCATTTGGAAAAGGAATATCTTTGCCGAGCAA 713
 QY 186 GlnLeuValAspCysAlaSerAlaPheAsnAsnGlyCys--GlyGlyLeuProSer 204
 DB 712 CAGCTGTGGATGTGCTGCTACTTTCAATTAATTTGGTGTCTCATGGTGGACTTCTCTCT 653
 QY 205 GlnAlaPheGluTyrrValLysTyrrAsnGlyGlyIleAspThrGluGlnThrTyrrProTyrr 224
 DB 652 CAAGCTTTGATACATAATAATACAAACGGTGGCTCGACACGGAGGAGGCTTATCCTTAC 593
 QY 225 LeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyValLysValIleAsp 244
 DB 592 ACCGAAAGAGGGGTGCGTCAAAATTTTCAGCGAAAACATCGGTGTACAAAGTCCGTGAC 533
 QY 245 SerIleAsnIleThrLeuGlyValAlaGluAspGluLeuLysHisAlaValGlyLeuValArg 264
 DB 532 TCTGTCAACATTAACCTGGTGCAGAGATGAATGAGACACGGGTGGGTGGTAAAG 473
 QY 265 ProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrrLysGlyValTyrr 284
 DB 472 CCAGTGTGGTGGCGTTTGGAGTTGTATCAATTTTATTAAGAGGGAGTGT 413
 QY 285 SerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGly 304
 DB 412 ACTAGCATATCATGTGGTAACACTCAATGATGTAAACCATGCAAGTGTAGCAGTGGT 353
 QY 305 TyrGlyValGluAspGlyIleProTyrrTrpLeuIleLysAsnSerTrpGlyThrAsnTrp 324

DB 352 TATGGAGTTGAAGACGACGTCCTGCTACTGCTTATAAAGACTCATGGGAGGTGAATGG 293
 QY 325 GlyAspAsnGlyTyrrPheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCys 344
 DB 292 GGAGACAAATGGCTACTTCAAGATGGAAATGGGAAGAACAATGCTGTGTTGCAACATGT 233
 QY 345 AlaSerTyrrProIleValAla 351
 DB 232 TCATCGTATCCGTTGTAGCC 212
 RESULT 10
 AAC40628
 ID AAC40628 standard; DNA; 1363 BP.
 AC AAC40628;
 XX 17-OCT-2000 (first entry)
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 28994.
 DE Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 OS
 PN EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-00301439.
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 03-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 13-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0130891P.
 PR 30-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 17-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144325P.
PR 18-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145119P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147152P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150586P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157855P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 4,12e-131
Score: 1259.50
Percent Similarity: 77.44%
Best Local Similarity: 66.57%
Length: 1363
Matches: 239
Conservative: 39
Mismatch: 73

Query Match:	67.53%	Indels:	8
DB:	3	Gaps:	4
US-10-087-714-2 (1-352) x AAC40628 (1-1363)			
Qy	1	MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer	17
Db	82	ATGCTCGGAACAACCTATCATAGTAGTTTGGTGGTCTCGCGCGCATGACGA	141
Qy	18	ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp	37
Db	142	CGCGCTAATATCGGATTCGATAGTCAACCCGATCCGAATGGTCTCCGATGGTCTCCGG	201
Qy	38	SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla	57
Db	202	GAGTAGAGAAATCGTTCACGATCTTAGTCAATCTGTCAGTCTCTCTTTCTGCT	261
Qy	58	ArgPheAlaArgAArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPhe	77
Db	262	CGCTTCACTCACCGATATGGTAAAGATATCAGAACGTGGAGGAGATGAAGCTTCGATT	321
Qy	78	GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr	97
Db	322	TCGATTTTCAAGGAGAACTTCGATTTGATCAGATCCACCAACAAGAGCGTTATCTTAC	381
Qy	98	ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTyrGluGluPheArgThrAsnArgLeu	117
Db	382	AAACTCGGTGTTAATCAATTTGCTGATTTGACATGGCAAGAGTTTCAAAGGACCAAGCT	441
Qy	118	GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal	137
Db	442	GGTGCTCTCAGAACTGCTCGCCACTTTAAAGGCGCAGCCACAAGGTCACAGAAGCAGCT	501
Qy	138	LeuProValThrArgAspTyrArgGluGlnGlyIleValSerProValLysAspGlnGly	157
Db	502	CTTCCTGAAACAAAGAACTCGAGAGAAGATGGTATCGTTAGTCGGTCAAGATCAGGGA	561
Qy	158	SerCysGlySer---TrrThrPheSerThrThrGlyAlaLeuGluAlaIleTyrThrGln	176
Db	562	GGTGTGGATCTTCCTCGACATTGAGCAACTGAGCTCTTGAAGCAGCTTACCATCAG	621
Qy	177	LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe	194
Db	622	GCATTTGGAAGGAATATCTCTCTGAGCAACAGCTTGTGGATTGTGCTGGAGCTTTC	681
Qy	195	AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn	213
Db	682	AATAACTATGTTGCAATGTGGCTCTCTTCTCAAGCCTTTGAATACATAAATCCAAC	741
Qy	214	GlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhe	233
Db	742	GGTGGCCTCGACACAGAGAAAGCTTATCTTATACCGGTAAAGATGAACCTGCAATTT	801
Qy	234	LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGlu	253
Db	802	TCAGCTGAAACAGTGTGTGTACAAGTCTCAACTCAGTCAACATTACTCTGGGTGCTGAA	861
Qy	254	AspGluLeuLysHisAlaVal-GlyLeuValArgProValSerValAlaPheGluValVa	273
Db	862	GATGAATGTAAGCATGCGGTTGGBATTTGGTACGGCCAGTAGCATAGCATTTGAGGTTAT	921
Qy	273	LysGlyPheAsnLeuTyrIlysGlyValTyrSerSerAspThrCysGlyArgAspPr	293
Db	922	ACACTCGTTCGGCTTTTACAAGAGTGGAGTTTACACTGATAGTCACTGTGGAAGTATC	981
Qy	293	oMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTy	313
Db	982	AATGGATGTGAACCAACCGCGGTTTGGCGGTTTGGTTATGGAGTTGAGAGCGGTGAC	1041
Qy	313	rTrrLeuIleLysAsnSerTrrGlyThrAsnTrrGlyAspAsnGlyTyrPheLysMetGl	333
Db	1042	TTGCTTATTAAAGAACTCATCGGAGCGGATTTGGGGGACAAAGGTTACTTCAAGATGGA	1101

```
Db 20 TCACGCTCTCGCTCGTATTGTTCTCGTCGCCGCCCTTTTCGCTACTGCACTTGCCTG 79
Qy -----PheGluGluAspAsnProIleArgSerValThrGlnArgProAspSerIle 39
Db 80 CGCGGCACCTTCGCTGATAAGAAATCCGATCAGCAAGTCGTA---TTTCCTCGATGAGCTG 136
Qy 40 GluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPhe 59
Db 137 GAGACGGGATTCCTCAAGTCGTGGCCAGACTCGCAGTGTCTCTCTCTCGTCGCTTT 196
Qy 60 AlaArgArgTyrGlySerTyrGlySerGluGluGluLeuIleGlyLeuArgPheGlyIle 79
Db 197 GCTATCAGGCATCGGAAAGGATGACTCCGTGTGAAGAGATCAACAAAGGTTTGAGATA 256
Qy 80 PheValGluAsnLeuAlaPheIleArgSerThrAsnArgGlyAspLeuSerTyrThrLeu 99
Db 257 TTTTGGACAACTGAGATGATCCGATCGATACAGAAAGGACTATCATCAAACTC 316
Qy 100 GlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAla 119
Db 317 GGTATCAATGAGTTTACCACCTTAACTGGGATGATTCGTTAAACACAAAGTTGGGGCA 376
Qy 120 AlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuPro 139
Db 377 TCTCAAACTGTCTGCCACCACAAAGGGAATCTCAGCTCACTAACTGCTTCTGCCA 436
Qy 140 ValThrArgAspTrpArgGluGlnGlyIleValSerProValGlyAspGlnGlySerCys 159
Db 437 GAGACGAAAGACTGGAGGAAGATGTTATTGTAGCCAGTGAAGCACAGGCGCAAGTGC 496
Qy 160 GlySer---TrpThrPheSerThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThr 178
Db 497 GGATCTGTGTCGACATTCAGCACTACTGTGTGACTAGAGCGACCTTATCCCAAGCATTT 556
Qy 179 GlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn 196
Db 557 GCGAAGGAATCTCTGTGCAGACGACGCTGTGTGCACTGTGTGCTGAGCTTTTAATAAC 616
Qy 197 PheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValGlySerGlyGly 215
Db 617 TTTGGTGTCAATGAGGGTGGCTCTCTCAAGCATTTTGAGTACATTAATTCATGTTGGT 676
Qy 216 IleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheGlyGln 235
Db 677 CTTGATATTGGTGTGACATATCCATACCGCGCAAGATGGCATATGTAATTCACAA 736
Qy 236 GluAsnValGlyValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGlu 255
Db 737 GCAATATTGGTGTCAAAGTCATCAGTTCTGTCAATATTACCTGGGTCTGAATATGAA 796
Qy 256 LeuIleHisAlaValGlyLeuValArgProValSerValAlaPheGluValGlyGly 275
Db 797 CTGAATACCGATGTGATTTGGTTAGCCCTGTTGCTGTTGCTTTTGGGTGTAAAGGG 856
Qy 276 PheAsnLeuTyrIleGlyValTyrSerSerAspThrCysGlyArgAspProMetAsp 295
Db 857 TTCAACAGTACAAAGAGCGAGTTTACGCCAGCAGCTGAATGTGGCGACATCCCATGGAC 916
Qy 296 ValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeu 315
Db 917 GTAAACCATGCTGTTCTGTGTGGGTACGGTTCGGTGTGAATGTTACTCCCTACTGGCTC 976
Qy 316 IleIleAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheIleMetGluLeuGly 335
Db 977 ATAAAGAACTATGGGGAGCAGATTTGGGGTGAGGATGATCTCAAAATGGAGATGGGA 1036
Qy 336 LysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 1037 AAGAAATATGTGTGTGTGAACCTTGTGCATCTCTACCAATCTGTGCC 1084
RESULT 12
ADB94877
ID ADB94877 standard; cDNA; 1652 BP.
```

```
XX ADB94877;
AC
XX
XX 04-DEC-2003 (first entry)
XX
XX Programmed cell death pathway cysteine protease cDNA #42.
XX
XX programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
XX DNase; lls; lsd1; nucellin-like aspartic protease; annexin; prohibitin;
XX fen-like protein; rac2; retinoblastoma-related protein; SINA; TF1D;
XX TEGT; xylogenic RNase; pur-alpha; cyteine protease; RPPs-like protein;
XX SP 91 NADPH oxidase subunit; NPS-like protein; BAG-1;
XX defender against cell death; lethal leaf spot; lesion stimulating death;
XX seven in absentia; transcription initiation factor;
XX testis enhanced gene transcript; gene; ss.
XX
XX Pinus radiata.
XX
XX US2003082724-A1.
XX
XX 01-MAY-2003.
XX
XX 14-AUG-2002; 2002US-00219220.
XX
XX 04-JUN-1999; 99US-00325932.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Flinn B, Lasham A;
XX
XX WPI; 2003-786916/74.
XX P-PSDB; ADB94919.
XX
XX New isolated polynucleotide useful for modulating programmed cell death,
XX altering the development cycle of plant cells, and subsequently modifying
XX plant development.
XX
XX Claim 1; Page 154; 214pp; English.
XX
XX The invention describes an isolated polynucleotide (I) comprising a
XX nucleotide sequence that is one of 145 fully defined sequences of 221-
XX 3415 base pairs (bp), given in the specification, and/or its complements,
XX reverse complements, reverse sequences, or sequences having 75, 90 or 95
XX % sequence identity to or that hybridise under stringent hybridisation
XX conditions to one of the 145 sequences. The methods and compositions of
XX the present invention to do with (I) are useful for modulating programmed
XX cell death and thereby altering the development cycle of plant cells, and
XX altering plant development. This sequence encodes a protein associated
XX with the programmed cell death pathway.
XX
XX SQ Sequence 1652 BP; 476 A; 318 C; 393 G; 465 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1,81e-118 Length: 1652
XX Score: 1148.00 Matches: 228
XX Percent Similarity: 75.21% Conservative: 42
XX Best Local Similarity: 63.51% Mismatches: 79
XX Query Match: 61.55% Indels: 10
XX DB: 9 Gaps: 8
XX
XX US-10-087-714-2 (1-352) x ADB94877 (1-1652)
Qy 1 MetAlaAlaLysLeuLeuPhePheLeu-----LeuPheLeuValSerAlaLeuSerVal 18
Db 264 ATGGCTCGCGCTCCGCCCATTTGTTCTGAGCAGCTTTGTTTTCGTG---GCCGTATCTGTA 320
Qy 19 AlaLeuAlaGly---PheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 321 TCTGCTGCGGGTCTGTTTGAAGAACTGGGTATATTGATATGTTGTCCTGACACAGATCCAA 380
Qy 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 381 GACCTGGAGTCCAAATTTTCGTTAAATCTTGGAAACCAATCCCAAGTCTCTACAATTTGCG 440
```


PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 8,226-114
Score: 1108.00
Percent Similarity: 45.22%
Best Local Similarity: 39.06%
Query Match: 59.41%
DB: 3

Length: 2149
Matches: 241
Conservative: 38
Mismatches: 72
Indels: 267
Gaps: 6

US-10-087-714-2 (1-352) x AAC44178 (1-2149)

Qy 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 81 ATGTCTGGCAAAACCAATCCTATCATCAGTAGTTTGGTGTTTCGTCCGCCATCAGCA 140
Qy 18 ValAlaLeuAlaGlyPheGlucIuAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 141 GCGGCTAATATCGGATTGCATGAGTCAAACCCGATCCGAATGGTCTCCGATGGTCTCCGG 200
Qy 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 201 GAGGTAGAAGAAATCTTTCCAGATCTTAGGTCAATCTCGTCACGTTCTCTCTTCGCT 260
Qy 58 ArgPheAlaArgGlyTyrglyLysSerTyrglySerGluGluGluIleValylsArgPhe 77
Db 261 CGCTTCACTCACCGATATGTAATAAGTATCAGAACGTCGAGGAGATGAAGCTTCGATTC 320
Qy 78 GlyIlePheValGluAsnLeu----- 84

Db	321	TCGATTTTCAAGGAGATCTTGATGGTAATGACTCCAACTTATTTGATAGTGTATTTATGTT	380
Qy	84	-----	84
Db	381	CAGATAATGCCGATGACTTTGTCTATGCTCAGCTCCACCGATTTTGAGAACACGACGACTT	440
Qy	84	-----	84
Db	441	CCGTCCAGCGTCCAGGCTGCTCCCTCAGATTTCAGGTTATGCGGCTCAATTCGCTGCGT	500
Qy	84	-----	84
Db	501	ATATCGCTTGCTGATTAGTGCAGCTTCCCTTCAGCGGGATTTCATACAGCGCCAGCC	560
Qy	84	-----	84
Db	561	ATCCGTATCATCATATACACAGTCAAAAGGTGACAGCAGGCTCATAGACGCCCCAGCGT	620
Qy	85	-----	86
Db	621	CGCCATAGTGGTTTCAACGAATACGTGCGCAACACCGTCTTCGGAGAGCTGTCTATACGC	680
Qy	86	-----	86
Db	681	GTAACACAGCCAGCGCTGGCGCATTTAGCCCGACATAGCCCCACTGTTTCGTCATTTC	740
Qy	86	-----	86
Db	741	CGCCACAGCATGACGTCACTGCCCGCTGTATGCCGCGAGTTACCGACTGCCGCTGAG	800
Qy	86	-----	86
Db	801	TTTTTAAGTACGTAAATCGTTGTGAGGCCAACGCCATAATGCCGGCTGTTCGCCGG	860
Qy	86	-----	86
Db	861	CATCCAAAGCCATTCATGGCCATATCAATGATTTCTGTGCTGCTACCGGGTTGAGAAGCG	920
Qy	86	-----	86
Db	921	GTGTAAGTGAATGCACTTGCCATGTTTACGGCAGTGAGCAGACAGATAGCGCTGATGT	980
Qy	86	-----	86
Db	981	CCGCGGTGCTTTTGGCTTACGCCACCCCGCTAGTAGCTGAACAGGAGGACAGCTG	1040
Qy	86	-----	86
Db	1041	ATAGAAACAGAGCCACTGGAGCACCTCAAAACACCATCATACATAATCAGTAAGTT	1100
Qy	87	-----	87
Db	1101	GGCAGCATCCAAATCTTGATTTGATCAGATCCACCAACAGAAAGCTTATCTTACAA	1160
Qy	99	LeuGlyIleAsnGlnPheAlaAspLeuThrTTPGluPheArgThrAsnArgLeuGly	118
Db	1161	CTCGGTGTTAATCAATTTGCTGATTGATTCGATGGCAGAGTTTCAAGGACCAAGCTTGGT	1220
Qy	119	AlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValaspGlyValLeu	138
Db	1221	GCTGCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCCACCAAGGTCACAGAACGAGCTCTT	1280
Qy	139	ProValThrArgaspTTPArgGluGlnGlyIleValSerProValIysaspGlnGlySer	158
Db	1281	CCTGAACAAAGACTGGAGAGAGATTGATCTGTGCTCCGTCAGAGATCAGGAGGT	1340
Qy	159	CysGlySer---TTPThrPheSerThrThrGlyAlaLeuGluAlaIatThrGlnLeu	177
Db	1341	TGTGATCTTCTGGACATTCAGCACGACTGGAGCTCTTCAGGACAGCTTACCATCAGGCA	1400
Qy	178	ThrGlySer-----ThrLeuSerGluGlnGlnLeuValaspCysAlaSerAlaPheAsn	195
Db	1401	TTTGGAAAGAAATATCTCTCTGTAGCAACAGCTTGTGGATTGTCTGGAGCTTTCAAT	1460

RESULT 14

AAC44535

ID AAC44535 standard; DNA; 1072 BP.

XX

AC AAC44535;

XX

DT 18-OCT-2000 (first entry)

XX

DE Zea mays DNA fragment SEQ ID NO: 43187.

XX

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway; metabolic; pathway;

KW promoter; termination sequence; corn; ss.

XX

OS Zea mays subsp. mays.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR

PR 05-MAR-1999; 99US-0123180P.

PR

PR 09-MAR-1999; 99US-0123548P.

PR

PR 23-MAR-1999; 99US-0125788P.

PR

PR 25-MAR-1999; 99US-0126264P.

PR

PR 29-MAR-1999; 99US-0126785P.

PR

PR 01-APR-1999; 99US-0127462P.

PR

PR 08-APR-1999; 99US-0128234P.

PR

PR 16-APR-1999; 99US-0129845P.

PR

PR 19-APR-1999; 99US-0130077P.

PR

PR 21-APR-1999; 99US-0130449P.

PR

PR 23-APR-1999; 99US-0130891P.

PR

PR 28-APR-1999; 99US-0131443P.

PR

PR 30-APR-1999; 99US-0132048P.

PR

PR 30-APR-1999; 99US-0132407P.

PR

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140921P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.

PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154775P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.

661 AAGAACTACTGGGGCGCTGACTGGGGTGATGAGGGTTACTTCAAGATGGAATGGGCANA 720

337 AsnMetCysGlyValAlaIaThrCysAlaSerTyrProIleValAla 351
:::|||||

721 AACATGTGGCGTGTGTACGTGTGCATCCTACCGCTGTGTGCA 765

RESULT 15
ACL19838/C
ID ACL19838 standard; DNA; 618 BP.
XX
AC ACL19838;
XX
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #9829.
XX
KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW gene; ss.
XX
OS Hordeum vulgare.
XX
FN WO2003057877-A1.
XX
PD 17-JUL-2003.
XX
PF 16-DEC-2002; 2002WO-IB005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403299.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SEP-2002; 2002JP-00327515.
XX
PA (UYN1-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda X, Kohara Y;
XX
DR WPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284pp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences
XX
SQ Sequence 618 BP; 142 A; 178 C; 152 G; 145 T; 0 U; 1 Other;
XX

Alignment Scores:
Pred. No.: 3,49e-85 Length: 618
Score: 846.00 Matches: 154
Percent Similarity: 87.13% Conservative: 22
Best Local Similarity: 76.24% Mismatches: 22
Query Match: 45.36% Indels: 4
DB: 8 Gaps: 3

US-10-087-714-2 (1-352) x ACL19838 (1-618)

Qy 154 LysAspGlnGlySerCysGlySer---TrrThrPheSerThrThrGlyAlaLeuGluAla 172
:|||||

Qb 616 AAAACCGGCGCCACTGGGGTCTCTGCTGGACGTCAGACACTTGGCGCACTTGGGCA 557
:|||||

QY 173 AlaTyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCys 190
Db 556 GCATATACTCAGGCCACTCGAAGAACATCTCTCTTCAGCAACAGCTGGTTGACTGT 497
QY 191 AlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyr 209
Db 496 GTGGTGGGTTCAATAACTTTGGATGCAACGAGAGCCTTCCATCCAGGCGTTTGAGTAC 437
QY 210 ValLysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGly 229
Db 436 ATCAATACAAATGGCGGGATCGACACCGAGAGAGTCTACCCCTTACAAGGGTGTCAATGGT 377
QY 230 IleCysAsnPhelYsGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThr 249
Db 376 GTCGTGCATTACAGCTGAAATGCTGCGGTTTCAGGTTTGGACTCAGTAAACATCACA 317
QY 250 LeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAla 269
Db 316 CTGAATGCCGAGGATGAAGACTGAAGAACGCCGTTGGGTTGCGCCAGTGAAGTGGGCC 257
QY 270 PheGluValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCys 289
Db 256 TTTCAGGTGATTACCGGTTTCAGGCAGTACAGAGCGGAGTTTACACGAGCCATTGT 197
QY 290 GlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAsp 309
Db 196 GGCACCTACCCCTCATGACGTGAACACCGCGCTTGGCAGTCGGCTATGCTGTGGAGAAC 137
QY 310 GlyIleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyr 329
Db 136 GCGGTCCCGTATTGGCTCATCAAGAACTCGTGGGGCGCCGACTGGGGTGACAAATGGTTAC 77
QY 330 PheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
Db 76 TTCAGATGGAATGGCAAGAACATGTGCGCTATTGCTACCTGGCGCTCTTACCCCTGTC 17
QY 350 ValAla 351
Db 16 GTGGCG 11

Search completed: March 29, 2004, 22:35:16
Job time : 383 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1300.5	69.7	1407	4	US-09-325-932A-107	Sequence 107, Appl
2	656.5	35.2	677	4	US-09-325-932A-111	Sequence 111, Appl
3	594	31.8	1600	4	US-09-325-932A-98	Sequence 98, Appl
4	590	31.6	1463	4	US-09-325-932A-92	Sequence 92, Appl
5	583	31.3	1390	3	US-08-821-994-61	Sequence 61, Appl
6	583	31.3	1441	3	US-08-821-994-63	Sequence 63, Appl
7	581	31.2	1434	3	US-08-821-994-62	Sequence 62, Appl
8	580	31.1	1428	4	US-09-325-932A-118	Sequence 118, Appl
9	573.5	30.8	1577	3	US-08-821-994-59	Sequence 59, Appl
10	567	30.4	1056	1	US-09-500-651-1	Sequence 1, Appl
11	567	30.4	1056	1	US-08-813-591-1	Sequence 1, Appl
12	562	30.1	1474	3	US-08-821-994-64	Sequence 64, Appl


```
/gene="At5g60360"
73. .1149
/gene="At5g60360"
/codon_start=1
/evidence=experimental
/product="putative cysteine proteinase AAPP"
/protein_id="AAN31819.1"
/db_xref="GI:23397068"
/translations="MSAKTILSGLVILVLAASAAANIGFDSNPTRMVSVDGLREVVEE
SVQILQGRHVSFAFTRHYKQYONVEEMKLRPSIFPENLDLIRNSTNKKGLSKYL
GVNQFADLTWQERFQTKGAQNCSTALAGSHKVTPEALPETKDMREDGIVSPVKDQG
GCSCSWFTTGALEAYHQARKGISLSEQLVDCAGAFNNGCNGGIPSOAFYIK
SNGLDTERAFYPTGDEKTCFAENVGQVLNSVNITLCAEDELKHAVLGVRPVSIA
FEVHSFRLYKSVYTDHSCGTFSDVNVHVLAVGYGVSDGVPPVYLIRKNSGADWDGDK
GYFKMKGKNCMCATCASYPVVA"
misc_difference 1128
/gene="At5g60360"
/notes="compared to genomic sequence"
/replace="a"
1150. .1409
/gene="At5g60360"
misc_difference 1393
/gene="At5g60360"
/notes="not present in genomic sequence"
ORIGIN
Alignment Scores:
Pred. No.: 6.29e-107 Length: 1409
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 8 Gaps: 4
US-10-087-714-2 (1-352) x BT000673 (1-1409)
```

```
Qy 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 73 ATGTCGCGAAACAAATCCATCATCAGTAGTTTGGTGGTCTCTCGTCGCGCATCAGCA 132
Qy 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 133 CGCGCTAATATCGGATTCGATGAGTCAACACCCGATCGAATGGTCTCCGATGCTCCCG 192
Qy 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 193 GAGTAGAGAGATCTGTTCCAGATCTTAGGTCATCTCGTCACGTTCTCTCTTTGCT 252
Qy 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluIleLysLysArgPhe 77
Db 253 CGCTTCACTCACCGATATGTTAAAGATATCAGAACCTGGAGGAGATGAAGCTTCGATTC 312
Qy 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
Db 313 TCGATTTCAGGAGATCTTGATTGATCAGATCCACCAACCAAGAGGCTTATCTTAC 372
Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 373 AAACCTCGGTGTAAATCAATTTGCTGATTGACATGGCAAGAGTTTCAAAGCACAAGCTT 432
Qy 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 433 GGTGCTGCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCCCAAGGTCACAGAGCAGCT 492
Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGly 157
Db 493 CTTCTGAAACAAAAGACTGGAGAGAGATGGTATCGTTAGTCCGGTCAAAGATCAGGGA 552
Qy 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaIleTyrThrGln 176
Db 553 GGTGTGATCTTCTGACATCTCAGCACAACTGAGGCTCTTGAGGAGGCTTACCATCAG 612
Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnLeuValAspCysAlaSerAlaPhe 194
```

Search completed: March 29, 2004, 23:30:32

Job time : 3309 secs

6

[illegible]

RESULT 2

RESULT 2
US-09-325-932A-111

US-09-325-932A-111
; Sequence 111. Application US/09325932A

Sequence III, Applicant No. 6451604

; FACNIC NO. 0431004
; GENERAL INFORMATION:

; GENERAL INFORMATION;
; APPLICANT: Flinn, Barry

APPLICANT: Lasham, Annette

```

/ TITLE OF INVENTION: Compositions affecting programmed cell
/ TITLE OF INVENTION: death and their use in the modification of forestry plant deve-
/ FILE REFERENCE: 1022
/ CURRENT APPLICATION NUMBER: US/09/325,932A
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 206
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 111
/ LENGTH: 677
/ TYPE: DNA

```

Alignment Scores:	
Pred. No.:	3-86a-70
Score:	656.50
Percent Similarity:	71.6%
Best Local Similarity:	57.9%
Query Match:	35.20%
DB:	4
Length:	577
Matches:	131
Conservative:	31
Mismatches:	49
Indels:	15
Gaps:	4

US-10-087-714-2 (1-352) x US-09-325-932A-111 (1-677)

[illegible]

RESULT 3

RESULT 3
US-09-325-932A-98

```
; Sequence 98, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 1600
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-98
Alignment Scores:
Pred. No.: 5,77e-62 Length: 1600
Score: 594.00 Matches: 147
Percent Similarity: 55.28% Conservative: 57
Best Local Similarity: 39.84% Mismatches: 121
Query Match: 31.85% Indels: 44
DB: 4 Gaps: 15
US-10-087-714-2 (1-352) x US-09-325-932A-98 (1-1600)
Qy 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGlu 24
Db 116 CTCGGCTGGCCCTGACGACGATCGCTCGCCCTCGACATGTCATCGTCAGCAT 175
Qy 25 GluAspAsnProIleArgSerValThrGln-----ArgProAspSerIleGlu 40
Db 176 CGGGCCACCGGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
Qy 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisAlaArgPheAla 60
Db 236 ATGCGCGTC-----TACGAGAGCTGGCTC 259
Qy 61 ArgArgTyGlyLysSerTyGlySerGluGluGluLeuLysLysArgPheGlyIlePhe 80
Db 260 GCCAAGCAGCGCAAGGCTCAACGCCCTGGCGGAGAGGAGGAGGAGGAGGAGGAG 319
Qy 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArg--LysAspLeuSerTyThrLeu 99
Db 320 AAGGACACCTCCGGTTCATCGACACACACGCCCGCGGACCGGACCTACCGGTC 379
Qy 100 GlyIleAsnGlnPheAlaAspLeuThrTyrGluGluPheArgThrAsnArgLeuGlyAla 119
Db 380 GGCCTCAACCATGTCGCGGACCTCACTAAGCAGAGGTACCGGTCTCATGTACCTGG 439
Qy 120 AlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArg-----Phe 133
Db 440 AGGATGATCGGTCGGCGCGGCTCGGAGGCGCCCGACGATCGGTACGCGCGTGGCC 499
Qy 134 ValAspGlyValLeuProValThrArgAspTyrArgGluGlnGlyIleValSerProVal 153
Db 500 GCCGGGAGGAGCTGCGCGGCTCGCTCATTTGGAGGAGGAGGAGGAGGAGGAGGAG 559
Qy 154 LysAspGlnGlySerCysGlySer---TyrThrPheSerThrThrGlyAlaLeuGluAla 172
Db 560 AAGGACGAGGAGCTCGGAGGTGTCGGCGGTCTCTACATGTCGTCTGTGGAGGGG 619
Qy 173 AlaTyThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCys 190
Db 620 ATAAACAGCTTGTCGCTGGTATTGATCTCTCTCGGAGGAGGAGGAGGAGGAGGAG 679
Qy 191 AlaserAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTy 209
Db 680 GATACATCCCAACAGAA---GGATGTAATGGCGGCTCATGATGATGATGATGATGAT 736
Qy 210 ValLysTyAsnGlyGlyIleAspThrGluGlnThrTyProTyThrLeuGlyValMetGly 229
```

```
Db 737 ATTATCAACACGAGGCAATTCATCCGAGGAAGATTATCCCTATAGAGCTGTAGATAGC 796
Qy 230 IleCysAsnPhelysGlnGluAsnValGlyValLysValLysValLysValLysVal 248
Db 797 ACTTGTGACCAATACAGGAAGAACCAAGGTTGTGACGATTGACGATTATGAAGATGT 856
Qy 249 ThrLeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerVal 268
Db 857 CCAGAAATGATGAGAAACCAATTCGAAAGGCGAGTTGCT---AATCAACCACTCAGTGTG 913
Qy 269 AlaPheGluVal---ValLysGlyPheAsnLeuTyLysGlyValTyLysGlyValTy 287
Db 914 GCCATTGAAGCAGGAGCGCGGAATTCAGTTTATGATTTCGGGTATATTTACTGTC--- 970
Qy 288 ThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyGlyVal 307
Db 971 AATGTGGG-----ACAGCTCTGGATCATGGGTTACTGCAGTCGGATATGGCACA 1021
Qy 308 GluAspGlyIleProTyTrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsn 327
Db 1022 GAAACCGGAGTTGATTACTGTGATGTGAAGAACTCATGGGCGGTAGCTGGGAGAGCAA 1081
Qy 328 GlyTyThrPheLysMet-----GluLeuGlyLysAsnMetCysGly 340
Db 1082 GGTATCATCAAAATGGCAGCAAAATGTGGCAATAGCCCACTGGCAAA-----TGTGT 1135
Qy 341 ValAlaThrCysAlaSerTyTrpIle 349
Db 1136 ATAGCAATGGAGGCGCTCTTACCCCATC 1162
RESULT 4
US-09-325-932A-92
; Sequence 92, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 92
; LENGTH: 1463
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-92
Alignment Scores:
Pred. No.: 1,53e-61 Length: 1463
Score: 590.00 Matches: 155
Percent Similarity: 57.45% Conservative: 57
Best Local Similarity: 42.01% Mismatches: 120
Query Match: 31.64% Indels: 37
DB: 4 Gaps: 16
US-10-087-714-2 (1-352) x US-09-325-932A-92 (1-1463)
Qy 7 PhePheLeuLeu-PheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAs 26
Db 18 TTCTTCTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 63
Qy 26 pAsnProIleArgSer-----ValThrGlnArgProAspSerIle-----GluPr 41
Db 64 -CATCATGGCGATCCCGCCACCGCCCGCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 122
Qy 41 AlaIleLeuGlyValLeuGlySerCysArgHisAlaPhe-HisPheAlaArgPheAla 61
Db 123 GCGGTGCTGGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 182
```



```
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 62
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Brassica napus
; US-08-821-994-62

Alignment Scores:
Pred. No.: 1434
Score: 581.00
Percent Similarity: 56.68%
Best Local Similarity: 38.24%
Query Match: 31.15%
DB: 3

US-10-087-714-2 (1-352) x US-08-821-994-62 (1-1434)

QY 4 LysLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAla---Gly 22
DB 57 AAACCTCTCTCTTACTTCTCTTATACGTC-----TTCTGTTCTGTTAGCTCCGGT 107
QY 23 PheGlu-----GluAspAsnProIleArgSerValThrGlnArgProAspSerile 39
DB 108 TATGAGTCCATCATCATGAGCAACCATCTCTCAGTCTCCATCTCAGCTTCGTTGGAGAAC 167
QY 40 GluProAlaLeuLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPhe 59
DB 168 GATGAAGAAGTGATATCCATC-----TACTTAAGATGG 200
QY 60 AlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLeu-----LysLys 75
DB 201 TCCTTGAGCAGCGGAAAGTAACAGCAACAGCAGCGTATTATCAACCAACAGCAAGAA 260
QY 76 ArgPheGlyLePheValGluAsnLeuAlaPheLe-----ArgSerThrAsnArgLys 93
DB 261 AGATTCAATATTTCAAAGACAACCTAAGATTTCATCATGATACACAAACGAGAAACAAG 320
QY 94 AspLeuSerTyrThrLeuGlyLeAsnGlnPheAlaAspLeuThrTyrGluGluPheArg 113
DB 321 AACGCTACTTACAAAGCTTGTCTTAACCATATTCGCTGATCTCACTAAGATGATACCGG 380
QY 114 ThrAsnArgLeuGlyAla-----AlaGlnAsnCysSerAlaThrAlaHisGly 129
DB 381 AGTTTATACCTCGGGCAAGAACCGAGCTGTCCGCGCATCTACTAAGCCCAAGAACGTT 440
QY 130 AsnHisArgPheValAspGlyVal-----LeuProValThrArgAspTyrArg 145
DB 441 AACATGAATAACTCAGCGCGAGTAACAGCCTGGAGGTTCCGGAGACGTTGACTGGAGA 500
QY 146 GluGlnGlyLeValSerProValLysAspGlnGlySerCysGlySer---TyrThrPhe 164
DB 501 AAGAAAGGAGCGGTTAATGCCATTAAAGACCAAGAACTTCCGGAAGTGTGTGGCGGTTT 560
QY 165 SerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----ThrLeu 182
DB 561 TCAACAGCTGCAGCAGTAGAAGGTATAAACAAGATCGTAAACAGAGAACCTCGTATCTTTG 620
QY 183 SerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGly 201
DB 621 TCCGACACAGAACTTGTGCACTGGACAAATCGTACAAACCA---GGCTGTAAACGGCGGT 677
QY 202 LeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLeAspThrGluGlnThr 221
DB 678 CTAATGGATTATGCTTTTCAATTATGATAAAACCGCGGATTAAACACCGAGAAAGAC 737

; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 62
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Brassica napus
; US-08-821-994-62

Alignment Scores:
Pred. No.: 1434
Score: 581.00
Percent Similarity: 56.68%
Best Local Similarity: 38.24%
Query Match: 31.15%
DB: 3

US-10-087-714-2 (1-352) x US-08-821-994-62 (1-1434)

QY 22 TyrProTyrLeuGlyValMetGlyLeCysAsnPhelysGlnGluAsnValGlyValLys 241
DB 738 TATCTTTACCACGGAACCAATGCAATGCAATCTTCTTACTTAAGATTTCAAGAGTTGTA 797
QY 242 ValIleAspSerile---AsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
DB 798 ACTATCGATGATACGAGATGTTCTCTAGTAAAGATGAAACCGGTTTGAAGAGCAGTT 857
QY 261 GlyLeuValArgProValSerValAlaPheGluVal---ValLysGlyPheAsnLeuTyr 279
DB 858 TCA---TACCAGCCTGTGAGTGTGCTATTGATGCTGCTGGAAGAGGTTTCCCAACATTAC 914
QY 280 LysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAla 299
DB 915 CAATCTGGAATCTTCACTGGA---AAGTGTGGT-----ACGAATATGGATCACGCT 962
QY 300 ValLeuAlaValGlyTyrGlyValGluAspGlyLeProTyrTyrLeuLysLysAsnSer 319
DB 963 GTGGTGGCGGTTGGTTATGGTTCAGAGACGCGGTTGACTATTGGATTGTACGTAACCTCT 1022
QY 320 TrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGly----- 335
DB 1023 TGGGTACACGTTGGGAGAGGATGTTATATTAGATGGAGAGAAACGTTGGCGTCTAAA 1082
QY 336 LysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
DB 1083 TCCGTAAGTCTGGGATTCGATAGAACCTCGTATCCGGTT 1124

RESULT 8
US-09-325-932A-118
; Sequence 118, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant deve
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 118
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Pinus radiata
; US-09-325-932A-118

Alignment Scores:
Pred. No.: 2,416-60
Score: 580.00
Percent Similarity: 53.41%
Best Local Similarity: 39.51%
Query Match: 31.10%
DB: 4

US-10-087-714-2 (1-352) x US-09-325-932A-118 (1-1428)

QY 2 AlaAlaLysLeuLeuPhePheLeuLeuValSerAlaLeuSerValAlaLeuAla 21
DB 19 TCCTCCACGCTGCTCTTCTTGGCTCTCTGATTTCTCTCTGTAATTTGGCTTCTTCGCGCC 78
QY 22 GlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAspSerileGluPro 41
DB 79 ATAAAGCCTGACGATGACCTTATTCGTAAGTACGAGCAAGTAGATTCA---GATCCA 135
QY 42 AlaIleLeuGlyValLeuGlySerCysArgHisAla---PheHisPheAlaArgPheAla 60
DB 136 CAGATCTCTGAGTCTCGGAGCGCCCTGTTCAACGCTGAAGCGCATTTTCAGCGCTTATC 195
QY 61 ArgArgTyrGlyLysSerTyrGlySerGluGluGluLeuLysLysArgPheGlyLePhe 80
```

196 AGGCGTACGGGAAGAAGTACTCGGGGCGGAGACGACGACGCGCTTCGGTGTCTTC 255
Qy 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerThrLeuGly 100
Db 256 AAGAGCAATTTACTAAGAGCCTTGGAGCAGAGAGCTCGACCCAGGCTCCCATGGC 315
Qy 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla 120
Db 316 GTCACAGAAATCTCTGAATTCACACAAGAGGAGTTCGACGACAGATATCTAGGCTCAGG 375
Qy 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140
Db 376 GCACACCGATCCGAGAGCCGACGATGCTCCAAATTTGGCCACAAAGATCTGCCGGAG 435
Qy 141 ThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGlnGlySerCysGly 160
Db 436 GAGTTCGATTCGAGAGAGAGGAGCCGTCACCGAGTTAAGATCAGGGATCGTGGGT 495
Qy 161 Ser---TrpThrPheSerThrThrGlyAlaLeuGluAlaIleThrGlnLeuThrGly 179
Db 496 TCCTGCTGGCTTTTCAGCACAAACGGGCGGTAGAGGGCGGAAATTTCTCTGAAGACGGG 555
Qy 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn----- 195
Db 556 AAGCTGGTGACCTTGAGGAGCAACAATTTGGTGAATCGGATCAGAGTGGATCTTCG 615
Qy 196 -----AsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyr 209
Db 616 GACGCAAGATCATGTGATCTGTTGCAATGCTGGCTAATGACGAGTGTCTTATCAATAT 675
Qy 210 ValLysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGly 229
Db 676 GCTCTGAAGCTGGTGGATGGAGGAAGAGGACTACCCATATCTCGAAAGACGGA 735
Qy 230 IleCysAsnPheLysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThr 249
Db 736 ACTTGACGCTTTAACAAGAACAAATTTGGSCACAGGT-----TCGAATTTTCAGC 786
Qy 250 LeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArg-----ProValSer 267
Db 787 GTTGTCTTATTGATGAGATCAATTTGCTGCAAACTGCTGGAAGATGGACCTCTATCA 846
Qy 268 ValAlaPheGluValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAsp 287
Db 847 GTGGGAATCAATGTGCA---TTTATGCAGACATACGTAGGAGGTGTATCTTSCCCATAC 903
Qy 288 ThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGly--- 306
Db 904 ATCTGCAAGACGCGATGTTGGAT-----CATGGTGTGCTCTCTGGTATGATGTTCT 957
Qy 307 -----ValGluAspGlyIleProTyrTrpLeuIleLysAsnSer 319
Db 958 CGAGCTTTGCTCCCATAGAAATAGGACAAA---CCCTACTGGATCATAAAGAACTCA 1014
Qy 320 TrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCys 339
Db 1015 TGGGACCTAACTGGGAGAAATGGATTCTACAAACTTTGACGGGGACATACGTTTGC 1074
Qy 340 GlyValAlaThrCysAlaSer 346
Db 1075 GGAATCAACAACATGGTTTCC 1095

RESULT 9

US-08-821-994-59
; Sequence 59, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A

; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 1577
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-59

Alignment Scores:

Pred. No.: 1,73e-59 Length: 1577
Score: 573.50 Matches: 141
Percent Similarity: 55.28% Conservative: 58
Best Local Similarity: 38.17% Mismatches: 114
Query Match: 30.75% Indels: 47
DB: 3 Gaps: 13

US-10-087-714-2 (1-352) x US-08-821-994-59 (1-1577)

Qy 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGlu 24
Db 63 CTCACCTTCTTCTCCCTTCTTCTAGTTCCTCTCTGAGCTTCTCATCATCATCTTCCGAT 122
Qy 25 GluAspAsnProIleArgSerValThrGlnArgProAspSerIleGluProAlaIleLeu 44
Db 123 GACATCTCCGAGCTG----- 137
Qy 45 GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgArgTyrGly 64
Db 138 -----TTGACGCTTGGTCGACGAGACACGGC 164
Qy 65 LysSerTyrGlySerGluGluGluIleLysLysArgPheGlyIlePheValGluAsnLeu 84
Db 165 AAAACGTCGCTTCGGAGGAAGAGACACACAGGATTCGAATCTTTAAAGACAAATCAC 224
Qy 85 AlaPheIle---ArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGln 103
Db 225 GACTTGTGTACACGACACAAACATCGCTAACTCTACTTCTCTCTCTCTCTCAATGCC 284
Qy 104 PheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCys 123
Db 285 TTCGGGATCTGACTCACCACGAGTTCAAGGCTCTCTCTTTGGAGGATCTCTGCTTCT 344
Qy 124 SerAla-----ThrAlaHisGlyAsnHis---ArgPheValAspGlyValLeuPro 139
Db 345 TCAGCTCCTTTGCTGATGGCTAAGGACACAGAGCTTTGAGAACGTTTCGGGGAAGGTTCCA 404
Qy 140 ValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCys 159
Db 405 GATCTGTTGATGGAGGAAGAAAGAGGCTGTATTAATGTCAAGATCAAGAAAGTGC 464
Qy 160 GlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaIleThrGlnLeuThr 178
Db 465 GGAGCGGTGTTGCTTCTTCGCGGACTGGAGCTATGGAAGGAATCAACAGATTTGTAACA 524
Qy 179 GlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsn 196
Db 525 GGAGATCTCATCGCTCTCTGAGCAAGAACTCATTTGATTTGTAAGTCATACACCAT 584
Qy 197 PheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGly 215
Db 585 ---GGATGCAATGGTGGTCTCATGGACTACGCTTTTCAATTTGTCTATTAATAACCATGG 641
Qy 216 IleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGln 235
Db 642 ATTGACACAGAAAGATTTATCTTTATCAAGAACGCTGATGGCACCTGTMAGAAAGATAG 701
Qy 236 GluAsnValGlyValLysValIleAspSer---IleAsnIleThrLeuGlyAlaGluAsp 254


```
Db 702 TTGAATAGAAAGTTGTGACAAATTGATAGCTAGCTAGCTGGTGTAAATCAATCAATGACGAGAA 761
Qy 255 GluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPhe---GluValVal 273
Db 762 GCGTTACTAGAAAGCTGTAGCG---GCTCAGCCAGTTAGTTGGTGTATCTGTGGGAGCGAG 818
Qy 274 LysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
Db 819 AGAGCGTTTCAGTTATCTACTCTAAGGAATATCTCT-----GGCCCATGTTCA 866
Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLeuProTyr 313
Db 867 ACATCATTCGACCCAGCAGCTGCTCATCTAGGATACGGTTCAAGAACGGTGTGATTAC 926
Qy 314 TrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheIysMetGlu 333
Db 927 TCGATCGTGAAGAAGCTCTTGGGGAAGAGTTGGGAATGGATGGGTTTATCCACATGCGAG 986
Qy 334 Leu-----GlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProfile 349
Db 987 CGTAACACCGGCAACGACGAGAGGATGTCGGGAATCAACATGCTGCTTCATATCCCATC 1046

RESULT 10
US-09-500-651-1
; Sequence 1, Application US/09500651
; GENERAL INFORMATION:
; APPLICANT: ASANO, MINAO
; APPLICANT: KAWAI, MITSUKO
; APPLICANT: MIWA, TETSUYA
; APPLICANT: NIO, NORIKI
; TITLE OF INVENTION: AMINOPEPTIDASE GX, AND A METHOD OF
; TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,651
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/813,591
; FILING DATE:
; APPLICATION DATA: JP 030458/1997
; FILING DATE: 14-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-845-0
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Glycine max
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: 1..1056
; OTHER INFORMATION: /note= "METHOD OF DETERMINING THE
; OTHER INFORMATION: CHARACTERISTICS: P"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..9
; OTHER INFORMATION: /note= "INSERTION SEQUENCE, METHOD
; OTHER INFORMATION: OF DETERMINING THE CHARACTERISTICS: E"
US-09-500-651-1
Alignment Scores:
Pred. No.: 5,63e-59 Length: 1056
Score: 567.00 Matches: 145
Percent Similarity: 56.65% Conservative: 51
Best Local Similarity: 41.91% Mismatches: 103
Query Match: 30.40% Indels: 47
DB: 16 Gaps: 16
US-10-087-714-2 (1-352) x US-09-500-651-1 (1-1056)
Qy 32 ValThrGlnArgProAspSerIleGluProAlaIleLeuGlyValLeuGlySerCysArg 51
Db 9 GTACGACGCGCCCA-----CGCGACAGGCGCGC 38
Qy 52 HisAla-----PheHisPheAlaArgPheAlaArg-ArgTyrG1 64
Db 39 CACGTTGCGCACCGAGGAGGAGCTGATGTCATGTACGAGCAGTGGTCTGTGAAGCAGCG 98
Qy 64 YLysSerTyrGlySerGluGluGluLeuLysLysArgPheGlyLlePheValGluAsnLe 84
Db 99 GAAGGTGTACACGCGCTCGCGAGGAGGAGGAGCGCTTCCAGATCTTCAAGAACACCT 158
Qy 84 uAlaPheIleArgSerThrAsn---ArgLysAspLeuSerTyrThrLeuGlyLleAsnG1 103
Db 159 GCGATTTCATCGACGACCACTCCGCGGAGGACCGCACTACCAAGCTCGGACTGAACCG 218
Qy 103 nPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla-----G1 121
Db 219 GTTCGCTGATCTCACCACGAGGATACAGGCGCAAGTACTTGGGAACCAAGATCATCC 278
Qy 121 nAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal-----Le 138
Db 279 CAACCGGAGCTCGAAGAACCCCGAGCAACCGCTACGCGCCAGCGTTCGGCGACAAAT 338
Qy 138 uProValThrArgAspTrpArgGluGluGlyLleValSerProValLysAspGlnGlySe 158
Db 339 GCCTGATTCGTTGATTGGAGGAAGGAGTGTCTCTCTCTGTCACCAAGACCAAGAGG 398
Qy 158 rCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLe 177
Db 399 CTGTGGGAGCTGTGGGCATCTCAGCAATCGGTGAGTAGAAGGATAAATAAGATAGT 458
Qy 177 uThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAs 195
Db 459 AACAGCGCAACTGATTTGCTTATCAGAACCAAGAAATGGTGGATGTGATACCTGGATATA 518
Qy 195 nAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnG1 214
Db 519 CCAA---GGATGCAATGGAGACTTATGGACTATGATTTGATTTTATATCAACATGG 575
Qy 214 YGlyIleAspThrGluGlnThrTyrTrpTyrLeuGlyValMetGlyLysCysAsnPhely 234
Db 576 CGGCATTGATTCGTAGAGGATTACCCATACCGTGGTGTGTGATGTAGTAGTGCACACATA 635
Qy 234 sGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGluAs 254
Db 636 TAGGAAAAATGCTAAAGTCTGTTTCTATGTATGATGCTACGAA---GATGTTCTCTCCCTATGA 692
Qy 254 pGlu-----LeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu-- 271
Db 693 TGAGTTAGCCTTGAAAAAGCCCGTTCGA---AATCAGCCCGTGGCGGTCTATTGAAGG 749
```

QY 272 -ValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyAr 291
Db 750 AGGGGAGGAGAAATTCATATATATGATGCTGCTGTTTACG -GGGAG 797
QY 291 qAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGly11 311
Db 798 ATGGGACACACACTAGATCATGTGTCGTGCTGTTGGGTATGGAACACACTAAAGTCA 857
QY 311 eProTyrTrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLy 331
Db 858 TGATTATTGGATCGTAAGGAATTCATGGGTTCTAGCTGGGAGAGGATGCTACATCAG 917
QY 331 sMetGluLeuGlyLysAsnMet-----CysGlyValAlaThrCy 344
Db 918 ATTAGAA-----AGAAATCTGTACACAGCATCAGGCAAGTGTGGAATTGCAATTGA 971
QY 344 sAlaSerTyrProLe 349
Db 972 GCCATCTTATCCCTT 987

RESULT 11
US-08-813-591-1
Sequence 1, Application US/08813591
Patent No. 5824534
GENERAL INFORMATION:
APPLICANT: ASANO, MINAO
APPLICANT: KAWAI, MISAKO
APPLICANT: MIWA, TETSUYA
APPLICANT: NIO, NORIKI
TITLE OF INVENTION: AMINOPEPTIDASE GX, AND A METHOD OF
TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,591
FILING DATE: 07-MAR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 051848/1996
FILING DATE: 08-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 030458/1997
FILING DATE: 14-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-845-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1056 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS

/ LOCATION: 1..1056 /note= "METHOD OF DETERMINING THE
/ OTHER INFORMATION: CHARACTERISTICS: P"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1..9
/ OTHER INFORMATION: /note= "INSERTION SEQUENCE, METHOD
/ OTHER INFORMATION: OF DETERMINING THE CHARACTERISTICS: E"
US-08-813-591-1
Alignment Scores:
Pred. No.: 5..63e-59 Length: 1056
Score: 567.00 Matches: 145
Percent Similarity: 56.6% Conservative: 51
Best Local Similarity: 44.9% Mismatches: 103
Query Match: 30.40% Indels: 47
DB: 1 Gaps: 16
US-10-087-714-2 (1-352) x US-08-813-591-1 (1-1056)
QY 32 ValThrGlnArgProAspSerIleGluProAlaIleLeuGlyValLeuGlySerCysArg 51
Db 9 GTACGACAGCGCCCA-----CGCGGACAAAGCGCGC 38
QY 52 HisAla-----PheHisPheAlaAArgPheAlaArg-ArgTyrGl 64
Db 39 CACGTTGCGCACCGAGGAGGAGCTGATGTCATGTACGAGCAGTGCCTCGTGAAGCAGG 98
QY 64 YLysSerTyrGlySerGluGluGluIleYsAlaArgPheGlyIlePheValGluAsnLe 84
Db 99 GAAGGTGTACAAACGCGCTCGCGGAGAGGAGCGCTTCAGATCTTCAAGGACAACT 158
QY 84 uAlaPheIleArgSerThrAsn---ArgLysAspLeuSerTyrThrLeuGlyIleAsnGl 103
Db 159 GCGATTTCATCGACGACCACTCCGCGGAGGAGCGAACCTCAAGCTCGGACTGAACCG 218
QY 103 nPheAlaIlePheLeuThrTrpGluPheArgThrAsnArgLeuGlyAlaAla-----Gl 121
Db 219 GTTCGCTGATCTCACCAAGAGGAAATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 278
QY 121 nAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal-----Le 138
Db 279 CAACCGGAGGCTCGGAAAGAGACCCCGAGCAACCGCTACGCGCCACGCTGCGGCGCAAAAT 338
QY 138 uProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySe 158
Db 339 GCCTGATTCGCTGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 398
QY 158 rCysGlySer---TrpThrPheSerThrTheGlyAlaLeuGluAlaAlaTyrThrGlnLe 177
Db 399 CTGTGGGAGCTGTGGGCAATCTCAGCAATCGGTGCGAGTAGAAGGAATAAATAAGATAGT 458
QY 177 uThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAs 195
Db 459 AACAGCGCAACTGATTTCGTTATCAGAACAGAAATTCGTTGATTCGTGATCTGATGATATA 518
QY 195 nAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGl 214
Db 519 CCAA---CGATGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 575
QY 214 yGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLy 234
Db 576 CGGCATTGATCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 635
QY 234 sGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGluAs 254
Db 636 TAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAA---GATGTTCCTCGCTATGA 692
QY 254 pGlu-----LeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu-- 271
Db 693 TGAGTTAGCCTTGAAAAAGGCGGTTGCA---AATCAGCCCGTGGCGCTGCTATTGAAGG 749
QY 272 -ValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyAr 291

```

Db      750 AGGGGCGAGGAATTCATATATATCTGGTGTATTCAGG-----GGGAG 797
Qy      291 GASPProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyI 311
Db      798 ATGTGCACAGCAGCTAGATCATGTGCTGCTGTGGTATGGAACGCTAAAGGTCA 857
Qy      311 eProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
Db      858 TGATATTGGATCGTAAGCAATTCATGGGTTCTAGCTGGGAGAGGATGGCTACATCAG 917
Qy      331 sMetGluLeuGlyLysAsnMet-----CysGlyValAlaLysThrCy 344
Db      918 ATTAGAA-----AGAAATCTTGCTAACACAGCATCAGCAAGTGTGGAATTGCAATTGA 971
Qy      344 sAlaSerTyrProile 349
Db      972 GCCATCTTATCCCTT 987

```

RESULT 12

```

US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643

```

GENERAL INFORMATION:

```

; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus

```

US-08-821-994-64

Alignment Scores:

```

Pred. No.:      3,86e-58      Length:      1474
Score:          562.00      Matches:      144
Percent Similarity: 55.41%      Conservative: 66
Best Local Similarity: 37.99%      Mismatches: 117
Query Match:      30.13%      Indels:      52
DB:              3          Gaps:      19

```

US-10-087-714-2 (1-352) x US-08-821-994-64 (1-1474)

```

Qy      2 AlaAlaLysLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAla 21
Db      42 TCACAAACTCATTTCTTACTTCTCTATACGTC-----GTCGTTTCATTAGCC 92
Qy      22 GlyPheGluGluAspAsnProileArgSerValThrGlnArgProAspSerileGluPro 41
Db      93 TCCGGTGATGATGCTCCATACCACTTCAATTAACCACTCATCTCAATCTCCATCCGAC-----143
Qy      42 AlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHis-----PheAla 57
Db      144 -----GGCTCATGGAGAACCGATGACAGAACTGAGTGGTCCATCTACTTA 185
Qy      58 ArgPheAlaArgArgTyrGlyLys-----SerTyrGlySerGluGluLys 74
Db      186 CAGTGGTGTGGGAGCAGCGGAAACTAGCAACAACACGATGTGTCACCAACCAAGAC 245
Qy      75 LysArgPheGlyIlePheValGluAsnLeuAlaPheIle-----ArgSerThrAsnArg 92
Db      246 GAAAGTTCAATATTTTCAAAGAACACCTAAGGTTTCATTGATCATCAACAATGAGAACAC 305

```

```

Qy      93 LysAspLeuSerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPhe 112
Db      306 AAGAACCGCTACTTACAAAGCTTGGTCTCACCATATTTCTCTCATCTCACTAAGCATGATGATC 365
Qy      113 ArgThrAsnArgLeuGlyAla-----AlaGlnAsnCysSerAlaThrAlaHis 128
Db      366 CGGAGGTATATACCTCGGGGCAAGAACCGAGTCTGTCCGCCCATCTACTAAGGCCNAGAAC 425
Qy      129 GlyAsnHisArgPheValAspGlyVal-----LeuProValThrArgAspTrp 144
Db      426 GTTAACATGAATACTACCGCGCGCAGTAAACAGCAGTGTGAGGTTCCGGAGACGGTTGATTGG 485
Qy      145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---TrpThr 163
Db      486 AGACGGAAGAGAGCGGTTAATGCCATTAAACCAAGAACTTCCGGAAGTTGTTGGCG 545
Qy      164 PheSerThrThrGlyAlaLeuGluAlaLysTyrThrGlnLeuThrGlySer-----Thr 181
Db      546 TTTTCGACAGCTGCAGCAGTAGAAGGTATAAACAGATCGTAACAGGAGAACTCATATCT 605
Qy      182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
Db      606 CTGTCGCAACAAGAACTTGTGCGACTGCGAGCAGATCTCTACAAACAA---GGCTGCAACGGT 662
Qy      201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
Db      663 GGTTAATGGACTATGCTTTTCAATTCATCATGAAAAACCGCGGTTTGAACACCGACAA 722
Qy      221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhelysGlnGluAsnValGlyVal 240
Db      723 GATTATCTTACCCTGCTGTTCCAAATGGAAATGCCAATCTTTTCTGAAAGAATTCAGAGTT 782
Qy      241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGlu-----LeuLysHis 258
Db      783 GTAACCTATTGATGTTTACGAA---GATGTTCTTACTGAAGATGAACGCGGTTTCAAGAGA 839
Qy      259 AlaValGlyLeuValArgProValSerValAlaPheGluVal---ValLysGlyPheAsn 277
Db      840 GCAGTTTCA---TACCAGCCCGTGGTGTGGCCATTGAAGCTGTGTGAAGAGTTTCCAA 896
Qy      278 LeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsn 297
Db      897 CATTACCAATCGGGATCTTCACCTGGA---AAGTGTGG-----ACAAATCTAGAT 944
Qy      298 HisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLys 317
Db      945 CATGCAGTGGTGGCTGTGGTTATGTTTTCAGAGAACCGGTATTGACTATTGCTATTGTAAGG 1004
Qy      318 AsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsn 337
Db      1005 AACTCGTGGGTACACGTTGGGAGAGAGGATGTTTACATTAGGATGGAG-----AGAAC 1058
Qy      338 Met-----CysGlyValAlaThrCysAlaSerTyrProile 349
Db      1059 TTGCAAGTCCAAAGTCCGGCAAGTGTGGAATTGCGGTTGAAGCTCGTACCCCGTT 1115

```

RESULT 13

```

US-08-821-994-82
; Sequence 82, Application US/08821994A
; Patent No. 6228643

```

GENERAL INFORMATION:

```

; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89

```

[illegible][illegible]

Db 894 GTGGCGTAAATCCGGTAAGTGTGGGATTGGGATAGAGACCTCGTATCCGGT 947

RESULT 15
US-08-821-994-60
; Sequence 60, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier R
; APPLICANT: Jenson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: EPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 60
; LENGTH: 1553
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-60

Alignment Scores:
Pred. No.: 1.69e-57 Length: 1553
Score: 557.00 Matches: 136
Percent Similarity: 54.87% Conservative: 61
Best Local Similarity: 37.88% Mismatches: 116
Query Match: 29.87% Indels: 46
Gaps: 13
DB:

US-10-087-714-2 (1-352) x US-08-821-994-60 (1-1553)
QY 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGlu 24
DB 74 ATCACTTCTCTCCCTCTCTCTAGTTTCTCTGAGCTTCCCATCATCATCTCCGAT 133
QY 25 GluAspAsnProLeuArgSerValThrGlnArgProAspSerLeuGluProAlaLeuLeu 44
DB 134 GACATCTCCGAGCTG-----TTTCACCGCTTGGTCCAGACACGGC 175
QY 45 GlyValLeuGlySerCysArgHisAlaPheHisAlaPheAlaArgPheAlaArgTyrGly 64
DB 149 -----TTTCACCGCTTGGTCCAGACACGGC 175
QY 65 LysSerTyrGlySerGluGluGluLeuLeuLysLysArgPheGlyPheValGluAsnLeu 84
DB 176 AAAACGTACGCTTCCGAGAGAGAGACACACAGGATTTGAAATCTTTAGACAAATCAC 235
QY 85 AlaPheLeu-----ArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGln 103
DB 236 GACTTCGTACACACACACACACCGCATCGTAACTTACTTACTTCTCTCTCACTCAATGCC 295
QY 104 PheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCys 123
DB 296 TTCCGGGATCTGACTCACCACGAGTTCAGGCGCTCTGCTCTTGGACTCTCTGCTTCTTCA 355
QY 124 Ser-----AlaThrAlaHisGlyAsnHis---ArgPheValAspGlyValLeuProVal 140
DB 356 GCTCGTGTCTGGTGGCTAAGGGAGAAAGTGTGAGAACGTTGGGGGCAAGATTCCAGAT 415
QY 141 ThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGly 160
DB 416 TCTGTGATTGGAGAGAGAGAGAGCTTTACTTAATGTCAAGATCAAGGAGCTGGGA 475
QY 161 Ser-----TrpThrPheSerThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGly 179
DB 476 GCGTGTCTGCTCTTCTCGCGGAGTGGAGCAATGGAGGAATCAACACAGATGTGTACAGGA 535
QY 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPhe 197

Alignment Scores:
Pred. No.: 9.83e-58 Length: 1102
Score: 557.00 Matches: 127
Percent Similarity: 59.43% Conservative: 62
Best Local Similarity: 39.94% Mismatches: 99
Query Match: 29.87% Indels: 30
Gaps: 14
DB:

US-10-087-714-2 (1-352) x US-08-821-994-86 (1-1102)
QY 56 PheAlaArgPheAlaArgTyrGlyLysSerTyrGlySerGluGluGluLeu----- 73
DB 12 TACTTAAGATGCTCTTGGAGCAGCGGAAAGTAAACAGACAGCAACGCGTATTATCAAC 71
QY 74 -----LysLysArgPheGlyLeuPheValGluAsnLeuAlaPheLeu-----ArgSer 89
DB 72 CAACAGACGAAAGTTCATTAATTTCAAGAGCAACCTAAGATTTCATGATCTACCAAC 131
QY 90 ThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrp 109
DB 132 GAGNACAACAGAACGCTACTTACAAGCTTGTCTAACCATATTGCTTAATCTCACTAAC 191
QY 110 GluGluPheArgThrAsnArgLeuGlyAla-----AlaGlnAsnCysSerAla 125
DB 192 GATCACTACCGGAGTTTATACCTCGGGGCAAGAACCGAGCCTGTCTCGCGCATCACTAAG 251
QY 126 ThrAlaHisGlyAsnHisArgPheVal-----AspGlyValLeuProValThr 141
DB 252 GCCAAGAGCTTAACATGAATACTCAGCGGCACTAAACGAGCTGGAGGTTCCGCTGAG 311
QY 142 ArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGlnGlySerCysGlySer 161
DB 312 GTTGAAGTGGAGACAGAAAGCGGTTAATGTCATTAAGACCAAGGAACTTTCGCGAAGT 371
QY 162 ---TrpThrPheSerThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer 180
DB 372 TGTGTGGCGGTTTCAACAGCTCGCAGCAGTAGAGAGGTATAAACAGATCGTACAGAGNA 431
QY 181 -----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGly 198
DB 432 CTCGTATCTTTGCGCAACAAGAACTTGTGCTGGCAGCAAAATCGTACAAACAA---GGC 488
QY 199 Cys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLeuAsp 217
DB 489 TGTAAAGCGCGCTTAATGATGTTTCAATTCATTAATGAAACCGCGGATTAAC 548
QY 218 ThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhelysGlnGluAsn 237
DB 549 ACCGAGAAAGACTATCTTACACGAGCAACCAATGGCAATGCAACTCTTTACTTAAGAT 608
QY 238 ValGlyValLysValLeuAspSerIle---AsnIleThrLeuGlyAlaGluAspGluLeu 256
DB 609 TCAAGAGTTTGAATCTGATGATAGTACGAAAGATGTTCTTAGTAAAGATGAAACCGGTTG 668
QY 257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluVal---ValLysGly 275
DB 669 AAGAGACAGTTTCA---TACCAGCCTGTGAGTGTGCTTATTCATGCTGCTGGAAGACT 725
QY 276 PheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAsp 295
DB 726 TTCAACATTAACCAATCTGAACTTCTCACTGGA---AAGTGTGCT-----ACGAAT 773
QY 296 ValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeu 315
DB 774 ATGATCACGCTGTGGTGGCGGTGTGTTATGGGTTCAGAGAACCGGCTGTGCTATTGGATT 833
QY 316 IleLysAsnSerTrpGlyThrAsnTrpGlyLysAsnGlyTyrPheLysMetGluLeuGly 335
DB 834 GTACGTAACCTCTTGGGCTACACGTTGGGAGAGAGGATGTTTACATTAGATGGAGAGAAC 893
QY 336 -----LysAsnMetCysGlyValAlaThrCysAlaSerTyrProLeu 349

```
Db 536 GATCTCATGCGCTCTCTGACGAGAACTAATTGATTGTACAAAGTCCTTACAAACGAT--- 592
Qy GlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIle 216
Db 593 GGATGCAATGGTGTCTCATGACGTACGCTTTTCAATTGTCATTAATAAACCATGGAATC 652
Qy 217 AspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhelLysGlnGlu 236
Db 653 GACACAGAGAAAGATTATCTCTTATCAAGAACGTGATGGCACCTGTAAAAAAGATAAGTTG 712
Qy 237 AsnValGlyValLysValIleAspSer---IleAsnIleThrLeuGlyAlaGluAspGlu 255
Db 713 AAAAGAAAGTTGTGACAAATTGATGCTGCGGTAAATCAACACGACGAGAAAGCG 772
Qy 256 LeuLysHisAlaValGlyLeuValArgProValSerValAlaPhe---GluValValLys 274
Db 773 TTA CTGGAAGCTGTAGCG---GCTCAGCCAGTTAGTGTGGCATCTGTGGCAGCGAGAGA 829
Qy 275 GlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspProMet 294
Db 830 GCGTTTCAGCTATACTCTAAGGGAATATTCTCT-----GCCCATGTTCAACA 877
Qy 295 AspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrp 314
Db 878 TCATTGGACCAACGCGAGTCTCATCGTAGGATACGGTTTCACAGAACGGTGTGATTACTGG 937
Qy 315 LeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeu 334
Db 938 ATCGTGAAGAACTCTTTGGGGAAGAGTTGGGGTATGGATGGGTTTATGCACATGCACCGT 997
Qy 335 -----GlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
Db 998 AACACCGGCAACTCGAAGGAGTATGTGGAATCAATATGCTCGCTTCGTATCCCATC 1054
```

Search completed: March 30, 2004, 00:11:56
Job time : 86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 23:30:35 ; Search time 330 Seconds
(without alignments)
3971.210 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAACKLFFLLFLVSALVAL.....ELGKNMGVATCASYPIVAV 352

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0			
Xgapop 10.0			
Xgapop 6.0			
Delop 6.0			

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10087714/runat_26032004_113454_1057/app.query.fasta_1.519
-DB=Published Applications NA -OEMT=fastap -SUFFIX=us-10-087-714-2 -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10087714@cgn_1_164@runat_26032004_113454_1057
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

	1	1819	97.5	1071	14	US-10-087-714-1
2	1321.5	70.9	1413	12	US-10-425-114-14419	Sequence 1, Appli
3	1320.5	70.8	1398	12	US-10-425-114-156	Sequence 14419, A
4	1319.5	70.8	1383	12	US-10-425-114-30654	Sequence 156, App
5	1319.5	70.8	1392	12	US-10-425-114-5115	Sequence 30654, A
6	1310.5	70.3	1367	12	US-10-425-114-5145	Sequence 5116, Ap
7	1310.5	69.7	1407	14	US-10-219-220-107	Sequence 5145, Ap
8	1280	68.6	1344	12	US-10-425-114-10837	Sequence 107, App
9	1280	68.6	1554	12	US-10-424-599-53151	Sequence 10837, A
10	1263	67.7	1282	9	US-09-770-445-20	Sequence 53151, A
11	1252	67.1	1183	12	US-10-425-114-2263	Sequence 20, Appl
12	1249	67.0	1390	12	US-10-425-114-18625	Sequence 2263, Ap
13	1234	66.2	1393	12	US-10-425-114-18625	Sequence 18625, A
14	1148	61.2	1652	14	US-10-424-599-122829	Sequence 122629, A
15	852.5	45.7	830	12	US-10-219-220-241	Sequence 241, App
16	845.5	45.3	888	14	US-10-425-114-3201	Sequence 3201, App
17	838	44.9	904	12	US-10-219-220-238	Sequence 238, App
18	828.5	44.4	813	12	US-10-425-114-6228	Sequence 6228, App
19	780.5	41.8	781	12	US-10-425-114-35711	Sequence 35711, A
20	759	40.7	1433	14	US-10-425-114-5160	Sequence 5160, Ap
21	759	40.7	1490	9	US-10-210-120-84	Sequence 84, Appl
22	759	40.7	1503	14	US-09-925-302-35	Sequence 35, Appl
23	759	40.7	1503	14	US-10-101-510-412	Sequence 412, App
24	691	37.1	1106	14	US-10-252-157-19	Sequence 19, Appl
25	656.5	35.2	677	14	US-10-101-510-153	Sequence 153, App
26	615	33.0	1392	9	US-10-219-220-111	Sequence 111, App
27	615	33.0	1392	11	US-09-938-842A-1372	Sequence 1372, Ap
28	609	32.7	1776	14	US-03-938-842A-1372	Sequence 1372, Ap
29	598.5	32.1	1410	12	US-10-219-220-236	Sequence 236, App
30	594.5	31.9	2054	14	US-10-425-114-6299	Sequence 6299, App
31	594	31.8	1600	14	US-10-219-220-234	Sequence 234, App
32	592.5	31.7	1730	12	US-10-219-220-98	Sequence 98, Appl
33	590.5	31.7	1746	12	US-10-425-114-32817	Sequence 32817, A
34	590.5	31.7	1746	12	US-10-425-114-30186	Sequence 30186, A
35	590.5	31.7	1746	12	US-10-425-114-22346	Sequence 22346, A
36	590	31.6	1463	14	US-10-425-114-30222	Sequence 30222, A
37	589	31.6	1392	15	US-10-219-220-92	Sequence 92, Appl
38	589	31.6	1712	12	US-10-618-644-6	Sequence 6, Appl
39	589	31.6	1857	12	US-10-425-114-10971	Sequence 10971, A
40	584	31.3	1407	14	US-10-425-114-29538	Sequence 29538, A
41	584	31.3	1568	12	US-10-253-165-723	Sequence 723, App
42	584	31.3	1576	12	US-10-425-114-5245	Sequence 5245, Ap
43	584	31.3	1576	12	US-10-425-114-1080	Sequence 1080, Ap
44	583.5	31.3	1631	12	US-10-425-114-26666	Sequence 26666, A
45	583.5	31.3	1570	12	US-10-425-114-33813	Sequence 33813, A
			1575	12	US-10-425-114-33200	Sequence 33200, A

ALIGNMENTS

RESULT 1

US-10-087-714-1
: Sequence 1, Application US/10087714
: Publication No. US20030070188A1
: GENERAL INFORMATION:
: APPLICANT: Havkin-Frenkel, Daphna
: APPLICANT: Podstolski, Andrzej
: APPLICANT: Dixon, Richard A.
: TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
: FILE REFERENCE: DMC10099
: CURRENT APPLICATION NUMBER: US/10/087,714
: CURRENT FILING DATE: 2002-02-28
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: 09/462,576
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: PCT/US98/14895
: PRIOR FILING DATE: 1998-07-15
: PRIOR APPLICATION NUMBER: 60/052,604
: PRIOR FILING DATE: 1997-07-15
: PRIOR APPLICATION NUMBER: 60/272,415
: PRIOR FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: Patent in version 3.1
: SEQ ID NO 1
: LENGTH: 1071


```

Db 436 GCGCAGCCAGAACTGCTCGCCACGCTTACCGGCAACACCGGATGACGCGCGCGCC 495
Qy 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLeuAsp 155
Db 496 GTTCGGCTGCCGAGAGCAAGAGCTGGAGGAGATGGGATTTGAGCCCGGAGAAAC 555
Qy 156 GlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174
Db 556 CAGGCGCACCTGTGATCATGTGACCTTACGACACTACTGCGCGACCTTGAGGTGCATAT 615
Qy 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
Db 616 ACCCAGGCAACTGGCAAGCCCATCTCTCTGAGCAACAGACTTGTGTACTGTGGTTT 675
Qy 193 AlaPheAsnAsnGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVallys 211
Db 676 GCATTCAACAATTTCCGATGCAACGGAGGCTTCCATCCCGGCTTTGATACATCAAA 735
Qy 212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCys 231
Db 736 TACATGTGGCTTGCACCTGAGCAATCTTACCCCTTACCAAGGTGTCAATGGAATCTGC 795
Qy 232 AsnPhelysGlnGluAsnValGlyVallysValleAspSerTleAsnIleThrLeuGly 251
Db 796 AAGTTTAAAGATGAAATGTTGGAGTCAAGTTTGGACTCGGTAAACACCCCTGGGT 855
Qy 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
Db 856 GCTCAGGATGAATGAAGGATGCTGTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 915
Qy 272 ValVallysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArg 291
Db 916 GTGATCACTGTTTTCAGCTGTACAGAGCGGAGTTTACACTAGCGACCAATGTGGAAT 975
Qy 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLe 311
Db 976 ACACCGATGATGGAACACGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1035
Qy 312 ProTyrTrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
Db 1036 CCTTACTGGCTTCATCAAGAACTCATGGGGCGCTGACTGGGGTGATGATGTTACTTCAAG 1095
Qy 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 1096 ATGGAAATGGGCAAGACATGTGGGTGTTGCTACGTGTGCTATCCCTATTTGCGCA 1155

```

RESULT 3

```

US-10-425-114-156
; Sequence 156, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 156
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700048693_FLI
US-10-425-114-156

```

Alignment Scores:

```

Pred. No.: 3,44e-163 Length: 1398
Score: 1320.50 Matches: 258
Percent Similarity: 78.8% Conservative: 26
Best Local Similarity: 71.6% Mismatches: 67
Query Match: 70.8% Indels: 9
DB: 12 Gaps: 6

US-10-087-714-2 (1-352) x US-10-425-114-156 (1-1398)
Qy 1 MetAlaAlaLysLeuLeuPhePheLeuLeuValSerAlaLeuSerValAlaLeu 20
Db 68 ATGGGCCCCAGCGCGCTGCTGCTCTCGCGCTGCTGCGCCACCGCGCGCGCG 127
Qy 21 Ala-----GlyPheGluGluAspAsnProIleArgSerValThrGluArgProAspSer 38
Db 128 GCCAACTCCGGCTTCGCGGACTCCAAACCGATCCCGCCGCTCACGAGCGCGCGCTCC 187
Qy 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisAla 57
Db 188 GGCCTCGAGTCCACCGCTTCGCGCGCTCGCGCCACCGCGGACGCGCTTCGCTTCGCA 247
Qy 58 ArgPheAlaArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPhe 77
Db 248 CGCTTCGCGCTCAGGTACGGCAAGAGTACGAGAGCGCGCGGAGGTCCACAAAGCGGTT 307
Qy 78 GlyIlePheValGluAsnLeuAlaPheLeuArgSerThrAsnArgLysAspLeuSerTyr 97
Db 308 AGGATCTTCTCGAGAGGCTCCAGCTGCTCGCTCCACCAACCGAAAGGCTCTCTCTAC 367
Qy 98 ThrLeuGlyLysAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 368 CGCTTCGCGCATCAACCGCTTCGCGCATGAGTGGGAGGAGTTCGCTGCGACCGCGCTC 427
Qy 118 GlyAlaAlaGluAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
Db 428 GCGCAGCGCCAGAACTGCTCGCGCACGCTTACCGCAACACCGGATCGCGCGCGCGCC 487
Qy 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyLysValSerProValLysAsp 155
Db 488 GTTGGCTGCCGAGACGAAAGACTGGAGGAGGATGGATGGAGTGGAGCCAGTGAATAAC 547
Qy 156 GlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174
Db 548 CAGGCGCACTGTGGATCATGCTGGACCTTCAGCACTACTGTGTCATCTTGAGGCTGCATAT 607
Qy 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
Db 608 ACCAGGCAACTGGCAAGCCCATCTCTCTGAGCAACACGCTTGTGACTGTGGTTT 667
Qy 193 AlaPheAsnAsnGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVallys 211
Db 668 GCATTCAACAATTTCCGATGCAACGGAGGCTTCCATCCAGGCGCTTTGAAATACATCAA 727
Qy 212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCys 231
Db 728 TACAAATGGTGGCTTGCACACTGAGGAATCTTACCCTTACCAGGAGTCAATGGAATTCG 787
Qy 232 AsnPhelysGlnGluAsnValGlyVallysValleAspSerIleAsnIleThrLeuGly 251
Db 788 AAGTTTAAAGATGAAATGTTGGAGTCAAGTTTGGACTCGGTAAACATCACCTCGGTT 847
Qy 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
Db 848 GCTGAGGATGAATCAAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 907
Qy 272 ValVallysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArg 291
Db 908 GTGATCACTGGTTTCAGGCTGTACAGAGCGGAGTTTACACTAGGACCAATGTGGAAT 967
Qy 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLe 311
Db 968 ACACCGATGATGGAACCAACGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027

```


QY 193 AlaPheAsnAspGlyCys---GlyGlyLeuProSerGlnAlaPheGlyTyrVallys 211
 DB 541 GCAATCAAAATTTCCGATGCAACAGGAGGCTTCATCCAGGCTTTGAATACATCAA 600
 QY 212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCys 231
 DB 601 TACAATGGTGGCTTGACACTGAGGAATCTTACCTTACCAAGGTGTCAATGGAATCTGC 660
 QY 232 AsnPheLysGlnGluAsnValGlyVallysValIleAspSerIleAsnIleThrIleuGly 251
 DB 661 AAGTTTAAAGATGAGAATGTTGGAGTCAAGGTTTTGGAGTCCGGTTAAACATCACCTGGGT 720
 QY 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 DB 721 GTGAGGATGAATGAGAGATGCTGTGGTCTGGTTCGCCAGTAGGTTGCCCTCGAG 780
 QY 272 ValValLysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArg 291
 DB 781 GTGACTCTGGTTTCAGGCTGTACAGAGCGGAGTTTACCTAGCGACCATTTGTGGAAT 840
 QY 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLeu 311
 DB 841 ACACCGATGGATGGAACACGCTGTTCTGGCTGTGGCTACGGTGTGGAAGATGGTGTG 900
 QY 312 ProTyrTyrLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
 DB 901 CCTACTGGCTCATCAAGAACTCATGGGCGCTGACTGGGGCGATGAGGTTACTTCAAG 960
 QY 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
 DB 961 ATGGAAATGGGCAAGAAATGTCGGGTGTGTACTGTGATCTACCTATCTATGTGCA 1020

RESULT 7

US-10-219-220-107
 ; Sequence 107, Application US/10219220
 ; Publication No. US20030082724A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Flinn, Barry
 ; TITLE OF INVENTION: Compositions affecting programmed cell
 ; TITLE OF INVENTION: death and their use in the modification of plant development
 ; FILE REFERENCE: 11000.1022c1
 ; CURRENT APPLICATION NUMBER: US/10/219,220
 ; CURRENT FILING DATE: 2002-08-14
 ; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
 ; PRIOR FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 290
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 107
 ; LENGTH: 1407
 ; TYPE: DNA
 ; ORGANISM: Eucalyptus grandis
 US-10-219-220-107

Alignment Scores:

Pred. No.: 1496-160 Length: 1407
 Score: 1300.50 Matches: 247
 Percent Similarity: 79.77% Conservative: 33
 Best Local Similarity: 70.37% Mismatches: 60
 Query Match: 69.73% Indels: 11
 DB: 14 Gaps: 5

US-10-087-714-2 (1-352) x US-10-219-220-107 (1-1407)

QY 9 LeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAspAsnPro 28
 DB 91 CTCCTCTCGTGGCGCTGCTGCTCTCCGCCGCGGCTGAGGCTTCGAGAGTCCAAACCCC 150
 QY 29 IleArgSerValThrGlnArgProAsp-----SerIleGluProAlaIleLeu 44
 DB 151 ATCCGCTCTTC-----CCCGACGGCGGCTCCGCGACCTCGAGTCTCTCCATCTGC 201

QY 45 GlyValLeuGlySerCysArgHisAlaPheHisAlaPheAlaArgPheArgTyrGly 64
 DB 202 CAGATCGTGGCGCGCACCGCCACGCGCTTCTCTTCCGCCGCTTCGCCAACAGGTATGG 261
 QY 65 LysSerTyrGlySerGluGluGluLysLysArgPheGlyIlePheValGluAsnLeu 84
 DB 262 AAGAGGTACGAGACCGCGGAGGAGATCAAGCTCGCGTTCGAGATCTTCAGGAGATCTC 321
 QY 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104
 DB 322 AAGTTGATCGATCCACCAACAGAGGCTTGCCTACACCTCGTCTCAATAGTTT 381
 QY 105 AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
 DB 382 GCTGATTGAGCTGGGAGGAGTTTCAAGAGGCACAGACTGGAGCTCTCAAACTCTCT 441
 QY 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
 DB 442 GCACACCAACAGGCGCACCAACAGCTCACGAGCAAGCTTCTCCCGAGATGAAGACTGG 501
 QY 145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---TrpThr 163
 DB 502 AGAGAAAAGGCGCATTTAAGCCCAATTAAGATCAGGGCGACTGTGATCTTGTCTGACT 561
 QY 164 PheSerThrThrGlyAlaLeuGluAlaIleTyrThrGlnLeuThrGlySer-----Thr 181
 DB 562 TTCACTACCACTGAGCTCTTGAAGTGTCTTATCACCAAGCATTCGGGAAACAAATCTCT 621
 QY 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
 DB 622 CTGCTGAGCAGCAGCTCGTGACCTGTGGGCTTTCAACAACATTTGGATGTAGTGGT 681
 QY 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
 DB 682 GGACTGCCATCCCAAGCCTTGAAGTACGTCAGCAAGCTGACGCTGCTTATACCGAGGAA 741
 QY 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
 DB 742 GCATATCTTATACCGCAGTGGATGAGTGTAGTGTCAATTTCTCGCTGATTAATGTGGTGC 801
 QY 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
 DB 802 CAAGTGTCTGACTCTGTATACATCACCTTGGGTGGTGGAGATGAATCAAGCATGCGATT 861
 QY 261 GlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
 DB 862 GCCTTCGTCGCGCAGTGGATGAGTGGCATTCAGGTCTGGAAGACTTCAGATTGTACAG 921
 QY 281 LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
 DB 922 TCGGGTGTCTACACGAGCGATACATCGGTAGCACTTCCATGATGTGAACCATGTCTGT 981
 QY 301 LeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSerTyr 320
 DB 982 CTCGCTGTGTTGATGAGTTGAAGATGGTGTTCCTGCTCTGCTCATCAGAAATTTCTGG 1041
 QY 321 GlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGly 340
 DB 1042 GGAGCAGACTGGGTGACCAACGATCTTCAAGATGGAGATGGAAAGAACATGTGTGGA 1101
 QY 341 ValAlaThrCysAlaSerTyrProIleValAla 351
 DB 1102 GTCGCTACTTGTGCATCATACCTCTGTGTGGCC 1134

RESULT 8

US-10-425-114-10837
 ; Sequence 10837, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E

```

; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 10837
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700980869_FLI
US-10-425-114-10837

```

Alignment Scores:	
Pred. No.:	6,966-158
Score:	1280.00
Percent Similarity:	77.87%
Best Local Similarity:	69.35%
Query Match:	68.63%
DB:	12
	12
	Gaps:
	5
	Indels:
	10
	Mismatches:
	69
	Conservative:
	34
	Matches:
	244
	Length:
	1344

UIS-10-087-714-2 (1-352) x UIS-10-425-114-10837 (1-1344)

Qy	1	MetaAlaLysLeuPheLeuPheLeuPheLeuValSerAlaLeuSerValAlaLeu	20
Db	28	ATGGCAGCGTGTCTCACTGCTCATTTTCGCGTTCGCGCGTGCCTGCGCGTGC	87
Qy	21	AlaGly-----PheGluGluAspAsnProIleArgSerValThrGlnArgProAspSer	38
Db	88	CGGGGCTCTCTTCGATCATGCCAACCTATCCGTTTAGCGTCG-----GAC	135
Qy	39	IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArg	58
Db	136	TTGGAGTCCGACGTCGTGGAGTGATCGGACAGAGCCGCCACGCTCTCTCTTCGCGCGC	195
Qy	59	PheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLeuLysLysArgPheGly	78
Db	196	TTTCGCTCGCGGCCACGGCAGCGGTACGTTCCGTCGACGAGATCCGCAACCGTTC	255
Qy	79	IlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThr	98
Db	256	ATCTTCTCTGATAATCTTAAACTCATCATGATCCACCAACAGGAGGTCTCTCACCTACAG	315
Qy	99	LeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGly	118
Db	316	CTCGGCGTCAACATCATTTTCTGCTACGTGGACTTGGGAGAGTTCCACGACCAACAGCTCGCG	375
Qy	119	AlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeu	138
Db	376	GCTCTCTCAGAATTGCTCTGCCAGCTCAAGGCAACACACAGGCTCACCGATGCTGTCTT	435
Qy	139	ProValThrArgAspTrpArgGluGlnGlnIleValSerProValLysAspGlnGlySer	158
Db	436	CCTGACGAGAAAGACTGGGAGAAAGAAAGGTATAGTCAGCCAAAGTTAAAGATCAAGCAAC	495
Qy	159	CysGlySer---TTrpThrPheSerThrThrGlyAlaLeuGluAlaIaTyrThrGlnLeu	177
Db	496	TGGGATCTTGCTGGACATTCAGCACACTGGTGCAATTGGAGGCGAGCTTACGCACAGGCC	555
Qy	178	ThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn	195
Db	556	TTTGGGAAGAATATCAGTCTTCTTCAGCAGCAGCTAGTGGACTGTGCGCGTGTCTTCAAT	615
Qy	196	AsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGly	214
Db	616	AACTTTGGCTGTATGGTGGTTGGCATCCCAAGCCTTTGTATACATAAATCAATAGT	675
Qy	215	GlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLys	234
Db	766	GGCTTCGACACAGAACCATATCCCTACACCGGAAAGATGGTGTCTGCAATTTACA	735

RESULT 9

```

US-10-424-599-53151
, Sequence 53151, Application US/10424599
, Publication No. US20040031072A1
, GENERAL INFORMATION:
, APPLICANT: La Rosa Thomas J
, APPLICANT: Kovalic David K
, APPLICANT: Zhou Yihua
, APPLICANT: Cao Yongwei
, TITLE OF INVENTION: Soy Nucleic Acid Mole
, TITLE OF INVENTION: Plants and Uses Ther
, FILE REFERENCE: 38-21(53223)B
, CURRENT APPLICATION NUMBER: US/10/424,599
, CURRENT FILING DATE: 2003-04-28
, NUMBER OF SEQ ID NOS: 285684
, SEQ ID NO 53151
, LENGTH: 1554
, TYPE: DNA
, ORGANISM: Glycine max
, FEATURE:
, OTHER INFORMATION: Clone ID: PAT_MRT3847

```

Alignment Scores:		
Pred. NO.:	8, 72e-158	Length:
Score:	1280.00	Matches:
Percent Similarity:	77.87%	Conservative:
Best Local Similarity:	68.35%	Mismatches:
Query Match:	68.63%	Indels:
DB:	12	Gaps:
		1554
		244
		34
		69
		10
		5

US-10-087-714-2 (1-352) x US-10-424-599-53151 (1-1554)

Qy	1	MetAlaAlaLysLeuLeuPhePheLeuPheLeuValSerAlaLeuSerValAlaLeu	20
Db	42	ATGGCAGCGTCTCAGTCTCATTTTCGGCTTCGCGCGTGCCTGCCTGC	101
Qy	21	AlaGly-----PheGluGluAspAsnProIleArgSerValThrGlnArgProAspSer	38
Db	102	GGCGGCTCTCTCTTCGATGAGCCACCTATCCGTTTACGCTCG-----GAC	149
Qy	39	IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisAlaArg	58
Db	150	TTGGATCGCAGGTCCTGGAGTGATCGGACAGAGCCACAGCTCTCTCTTCGCGCGC	209
Qy	59	PheAlaArgArgTyrGlyIysSerTyrGlySerGluGluIleLeuLysLysArgPheGly	78
Db	210	TTTCGTCGCGCGCAGCGGACGCTACCGTTCCGTCGACGAGATCCGACACCGCTTCGG	268

[illegible]

```

Db 372 ATATTCTGCAAAACCTCAGGTTCTACCTCCCAACAGAACCGCTTGCCCTACT 431
QY 99 LeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheThrAsnArgLeuGly 118
Db 432 CTCTCTGTTAATCATTTTCTGATGGACTTGGGAGGAGTTCAAAAGACACAGACTAGGC 491
QY 119 AlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeu 138
Db 492 GCTGCCCAAAATGCTCTCCCACTCTAACGGCAACCAACCACTCACTGATGCTGTTT 551
QY 139 ProValThrArgAspTrpArgGluGlnGlyIleValSerProValIleAspGlnGlySer 158
Db 552 CTTCCCAAGAGACTGGAGGAAGAGAAATAGTGAGTGATGTTAAAGATCAAGGCAGC 611
QY 159 CysGlySer---TrpThrPheSerThrGlyValAlaLeuGluAlaAlaTrpThrGlnLeu 177
Db 612 TCGGATCATGTGGACATTCAGCACTGAGCGCTTAGAAGCAGCCCTATGCACAGCA 671
QY 178 ThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn 195
Db 672 TTTGGGAAGAGATCTCTCTTTCTGACAGCAGCTAGTGATGTGCTGGCCCTTCAAC 731
QY 196 AsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTrpVal-LysTyAsnGln 214
Db 732 AACTTTGGCTGTAATGCTGGGTGTCATCACAAGCCTTTGAGTACATTTAAATAACAATGG 791
QY 214 YGlyIleAspThrGluGlnThrTyProTyLeuGlyValMetGlyIleCysAsnPheIle 234
Db 792 TGACTAGAGACAGAGAAAGCATATCCCTACACAGGAAAGATGGTGTCTGCAAAATTTCTC 851
QY 234 sGlnGluAsnValGlyValIleAspSerIleAsnIleThrLeuGlyValAlaGluAs 254
Db 852 AGCTGAAATGTTGCTGTTCAAGTCTCTGACCTGTGAATATCACCTTGGGTGCTGAGAA 911
QY 254 pGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVally 274
Db 912 TGAATTAACAACATGCAATGTCATTTCTTGGCGGCTAGTGTGGCTTTTCAGGTGGTGA 971
QY 274 sGlyPheAsnLeuTyLysLysGlyValTyLysSerSerAspThrCysGlyArgAspProme 294
Db 972 TGGTTTCCATTTCTACAGAAATGAGCTTTACACTAGTACATTTGGGTAGCATTCCCA 1031
QY 294 tAspValAsnHisAlaValLeuAlaValGlyTyGlyValGluAspGlyIleProTyTr 314
Db 1032 GGATGTCAACCATGCGCTCTGCTGCTGGGTATGAGATGGAATGGCGTCCCATATTG 1091
QY 314 pLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyPheLysMetGluLe 334
Db 1092 GCTCATAAAAATTCATGGGAGAAAGTTGGGTGAGAAATGGCTACTTCAAAATGGAAT 1151
QY 334 uGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyProIleValAla 351
Db 1152 GGGGAAGAACATGTGTGTTGTCACACTTGTGGCTTATCCAGTTGTGGCA 1203

RESULT 14
US-10-219-220-241
; Sequence 241, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE OF INVENTION: death and their use in the modification of plant development
; FILE REFERENCE: 11000.102261
; CURRENT APPLICATION NUMBER: US/10/219, 220
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 241
; LENGTH: 1652

```

```

; TYPE: DNA
; ORGANISM: Pinus radiata
US-10-220-241

Alignment Scores:
Pred. No.: 2,3e-140 Length: 1652
Score: 1148.00 Matches: 228
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 63.51% Mismatches: 79
Query Match: 61.55% Indels: 10
DB: 14 Gaps: 8

US-10-087-714-2 (1-352) x US-10-219-220-241 (1-1652)
QY 1 MetAlaAlaLysLeuLeuPhePheLeu-----LeuPheLeuValSerAlaLeuSerVal 18
Db 264 ATGCTCGCGTCTCGGCATTTCTGAGCAGCTTTGTTTGTCTG---GCGTATCTGTA 320
QY 19 AlaLeuAlaGly---PheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 321 TCTGCTCGCGGTCTGTTTGAAGAACTGGGTATATTGATATGCTCACTGCAAGATCCAA 380
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 381 GACCTGGAGTCCAAATTCGTTAAATCTTGGGAACCAATCCCAAGTCTCTACAAATTTGCG 440
QY 58 ArgPheAlaArgArgTyGlyLysSerTyGlySerGluGluGluIleLysLysArgPhe 77
Db 441 GAGTTCTGCTGCAGATATGCAAGAGGTACGATTTCTGTCATCAGCTTGTGCATAGATTTC 500
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTy 97
Db 501 AATCTCTTTGTGAAGAACGTTGGAGCTGATCGAGTCAAGAACAGAAATGCAAGTCTTCTTAT 560
QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 561 ACTTTGGCAATAATGAGTTTGTGACATAACATGCGAGGAATTCATGACAAATATTG 620
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 621 GGTCTCTTCAAGAACTGTTCTGGCTACC---CACAGTAAACCAATAGTTGACGTATGCCAG 677
QY 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGly 157
Db 678 CTTCCTCGGAAGAAAGACTGGAGACAGAGGCACTAGTCTGCTGTGAAAAACCAAGCC 737
QY 158 SerCysGlySer---TrpThrPheSerThrGlyAlaLeuGluAlaAlaTyThrGln 176
Db 738 CATTTGTGATCTCTGCTGCACATTGAGCACTACTGAGCACTAGAGGCTGCTTACTACTCAG 797
QY 177 LeuThrGlySerThr-----LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
Db 798 GCTACAGAAAGACTGTTATCTCTGTAACAGCAGCTGTTGTGACTGTGCTGGAGCATTT 857
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyValLysTyAsn 213
Db 858 AACAACTTTGTTGCAATGTTGGTGGACTGCCATCCCAAGCAATTTGAGTATATCAAGTACAAC 917
QY 214 GlyGlyIleAspThrGluGlnThrTyProTyLeuGlyValMetGlyIleCysAsnPhe 233
Db 918 GGAGGCTTGAACACTGAGGAAGCCCTATCTTACACCGCTAAAGATGGTGTGTTTGAATAT 977
QY 234 LysGlnGluAsnValGlyValIleAspSerIleAsnIleThrLeuGlyValAlaGlu 253
Db 978 GATGTAAACATATGTCGGTGTAAAGGTTGCTGATAGTGTCAACATCAGTTTGGGGGAGAA 1037
QY 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 1038 GATGACCTCAAAGCTGCTGCTGGGTCTAGTTGCGCTCTGCTAGTGGGCATTCCAAGTAAAT 1097
QY 274 LysGlyPheAsnLeuTyLysLysGlyValTyLysSerSerAspThrCysGlyArgAspPro 293
Db 1098 TCGGAATTTCCATTTTACAGGAGAGGTCTTTACAGCAACAGTGTGTGCTCAAGGCCCA 1157

```

QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrClyVal---GluAspGlyIlePro 312
Db 1158 ATGATGTCACCATGCTGTTTGGCCGTTGGTATGGTGTAGTGGAGGGGACTTCA 1217
QY 313 TyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMet 332
Db 1218 CACTGGATCATCAAGATTCTCTGGGGAAGAGCTGGGGTCTTGATGGATACCTTCAAGATG 1277
QY 333 GluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 1278 GAGTTAGGGAAGAAATATGTGTGGTGTGGACCTTGGCTTCGTACCTATTGTTTCT 1334

RESULT 15

US-10-425-114-3201
; Sequence 3201, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3201
; LENGTH: 830
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700243475_FLI
US-10-425-114-3201

Alignment Scores:
Pred. No.: 6,39e-102 Length: 830
Score: 852.50 Matches: 158
Percent Similarity: 89.06% Conservative: 13
Best Local Similarity: 82.29% Mismatches: 18
Query Match: 45,71% Indels: 3
DB: 12 Gaps: 2

US-10-087-714-2 (1-352) x US-10-425-114-3201 (1-830)

QY 163 ThrPheSerThrThrGlyAlaLeuGluAlaIleTyrThrGlnLeuThrGlySer----- 180
Db 4 ACCTTCAGCACTACTGGTGCACTTGAGGCTGCATATACCCAGCAACTGGCAAGCCCATC 63
QY 181 ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnGlyCys--- 199
Db 64 TCTCTCTGAGCAACAGCTTATTGACTGTGTTTGGATTCATCAATTCGATGCAAC 123
QY 200 GlyGlyLeuProSerGlnAlaPheGluTyrVallyTyrAsnGlyGlyIleAspThrGlu 219
Db 124 GGAGGCTTCCTCCAGGCTTGCATACATCAATCAATGTTGGCTTGACACTGAG 183
QY 220 GlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhelysGlnGluAsnValGly 239
Db 184 GAATCTTACCCCTTACCAGGTTGTCATGGAATCTGCAAGTTTAAAGATGAGAATGTTGA 243
QY 240 ValLeuValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAla 259
Db 244 GTCAGGTTTGGACTCGGTTTACATCACCTGGGTGCTGAGATGAACTGAAGATGCT 303
QY 260 ValGlyLeuValArgProValSerValAlaPheGluValVallyGlyPheAsnLeuTyr 279
Db 304 GTTGGTCTGGTTCGCCAGTTAGTGTGCTTCGAGGTGATCACTGGTTTCAGGCTGTAC 363
QY 280 LysLysGlyVallyTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAla 299

Db 364 AAGAGCGGAGTTTACCTAGCGACCATTTGTGAACTACCCGATGGATGTGAACACGCT 423
QY 300 ValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSer 319
Db 424 GTTCTGGCTGTTGGCTACGGGTGCGAAGATGGTGTACCTACTGGCTCATCAAGAACTCA 483
QY 320 TrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCys 339
Db 484 TGGGCGCTGACTTGGGGTGATGAGGGTTACTTCAAGATGGAATGGGCAAGACATGTGC 543
QY 340 GlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 544 GGTGTGTGTACGTGTGCATCCTACCTATTATTGTGCA 579

Search completed: March 30, 2004, 01:10:56
Job time : 342 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:03:58 ; Search time 33 Seconds
(without alignments)
1026.043 Million cell updates/sec

Title: US-10-087-714-2
Perfect score: 1865
Sequences: 1 MAAKLLFFLLFLVSALSVAL.....ELGKNMCGVATCASYPIVAV 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1325.5	70.9	360	2	S59598
2	1281.5	68.7	362	1	KHRZOG
3	1254.5	67.3	350	2	S71923
4	1252	67.1	361	1	KHBH
5	1249	67.0	356	2	S66348
6	1207.5	64.7	377	2	T47471
7	796	42.7	333	1	KRTH
8	763	40.9	335	1	KHHU
9	656.5	35.2	326	2	S43991
10	616	33.0	374	2	T03941
11	613.5	32.9	338	2	JCS443
12	609	32.7	462	2	JN0719
13	608	32.6	364	2	T12039
14	604.5	32.4	326	2	S53027
15	604	32.4	364	2	T46630
16	603.5	32.4	339	2	A53810
17	594.5	31.9	323	2	S19650
18	594.5	31.9	368	2	S47312
19	587	31.5	471	1	KHRZOB
20	586	31.4	322	2	S19649
21	585.5	31.4	455	2	T12041
22	584.5	31.3	344	2	JX0366
23	577.5	31.0	466	2	T06416
24	573	30.7	324	2	S47432
25	573	30.7	376	2	E85435
26	572.5	30.7	454	2	JC4848
27	571.5	30.6	326	2	T09259
28	570	30.6	328	2	JQ1121
29	566.5	30.4	331	2	JCS442

RESULT 1
S59598
cysteine proteinase (EC 3.4.22.-) 2 precursor [similarity] - maize
C/Species: Zea mays (maize)
C/Date: 15-Feb-1996 #Sequence_revision 01-Mar-1996 #text_change 20-Jun-2000
C/Accession: S59598
R/Domoto, C.; Watanabe, H.; Abe, M.; Abe, K.; Arai, S.
Biochim. Biophys. Acta 1263, 241-244, 1995
A/Title: Isolation and characterization of two distinct cDNA clones encoding corn seed
A/Reference number: S59597; MUID:96004895; PMID:7548211
A/Accession: S59598
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-360 <DOM>
A/Cross-references: EMBL:D45403; NID:G1688044; PIDN:BAA08245.1; PID:G644490
C/Superfamily: papain
C/Keywords: cysteine proteinase; hydrolase
F:167,307,321/Active site: Cys, His, Asn #status predicted

Query Match 70.9%; Score 1322.5; DB 2; Length 360;
Best Local Similarity 71.7%; Pred. No. 5.8e-103;
Matches 258; Conservative 28; Mismatches 65; Indels 9; Gaps 6;

QY	1	MAAKLLFFLLFLVSALSVAL--AGFEEDNPISVTORPDS--TEPAILGLVGSCHAFHFA	57
DB	1	MYPRRLFLVAVVLADTAAVVNSGFADSPFIPVTDRAASALESTVFAALGTRDALRFA	60
QY	58	RPARRYGKSYGSEEEIKKFGIFVENLAFIRSTNRKLSYTLGINQFADLTWEEFTNKL	117
DB	61	REAVRYGKSYESAEEVHKFRIFSESLSQLVRSNRRKLSYRLGINRFDMSWEFPRATRL	120
QY	118	GAQONCSATAHGNHRP--VDGVLVPTRDVREOGIVSPVKDQSCGS--WTFSTTGALEAY	174
DB	121	GAQONCSATLTGNHRRMAAAVALPETKDRREDIVSPVKNQGHGSCWTFSTTGALEAY	180
QY	175	TOLTGS--TLSEQLVDCASAFNNFC--GGLPSQAFYVYKNGGIDTEQTYPLGYMGIC	231
DB	181	TQATGPISLSQQLVDCGFANFCNGGLPSQAFYIKYNGGLDTEESYPVQGVNGIC	240
QY	232	NFKQENGVKVIDSINITGABDELKHVGLVRPVSVAFEVKGFNLYKKGVYSSFTGCR	291
DB	241	KFKENNVGVKLDVSNVITLGADELKDAVGLVRPVSVAFEVITGFRFLYKSGVYTSHCCT	300
QY	292	DPMDVNHAVLAVGYGVEDGIPYWLKNSMGTWNGDGYFKWELGKNMCGVATCASYPIVA	351
DB	301	TPMDVNHAVLAVGYGVEDGVPYWLKNSMGTWNGDGEYFKWENGMKNCVATCASYPIVA	360

RESULT 2
KHRZOG
oryzain (EC 3.4.22.-) gamma precursor - rice
C/Species: Oryza sativa (rice)

probable cysteine
cysteine proteinase
cysteine proteinase
cysteine L-like c
cathepsin L (EC 3.
cysteine proteinase
cysteine proteinase
actinidin (EC 3.4
cysteine proteinase
prestalk cathepsin
cathepsin L (EC 3.
cathepsin L (EC 3.
cysteine proteinase
cysteine proteinase
oryzain (EC 3.4.22
probable cysteine

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jun-2000
C:Accession: J00390; C40053
R;Watanabe, H.; Abe, K.; Emori, Y.; Hosoyama, H.; Arai, S.
submitted to JFID, May 1991
A;Reference number: J00388
A;Accession: J00390
A;Molecule type: mRNA
A;Residues: 1-362 <WAT1>
R;Watanabe, H.; Abe, K.; Emori, Y.; Hosoyama, H.; Arai, S.
J. Biol. Chem. 266, 16897-16902, 1991
A;Title: Molecular cloning and gibberellin-induced expression of multiple cysteine protease
A;Reference number: A40053; MUID:91358494; PMID:1885617
A;Accession: C40053
A;Molecule type: mRNA
A;Residues: 1-362 <WAT2>
A;Cross-references: GB:D90408; NID:G218184; PIDN:BAAL4404.1; PID:G218185
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; hydrolase; seed
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-144/Domain: propeptide #status predicted <PRO>
F;145-362/Product: oryzain gamma #status predicted <MAT>
F;128-258/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;166-209, 200-242, 300-350/Diulfide bonds: #status predicted
F;169, 309, 329/Active site: Cys, His, Asn #status predicted
Query Match 68.7%; Score 1281.5; DB 1; Length 362;
Best Local Similarity 66.9%; Pred. No. 1.6e-99;
Matches 242; Conservative 45; Mismatches 64; Indels 11; Gaps 6;
QY 1 MAAKLLFFLLFVLVLAAGFEEDNPIRSTQRPDS-IEPAILGVLSGRHAF 54
DB 1 MAARRILLLAAVAATSAVAASGFDSDNPIRSTQTHAASALESTVIALGRTGAL 60
QY 55 HFARFARYKSGSEIEIKRFGIFVENLAFIRSTRNRKDLSTYLGINQFADLTWEEFRT 114
DB 61 RFARFARVHGKRGDAEVRQRRIFSESLELVKSTNRRLGLPYRLGINFADMSWEEFOA 120
QY 115 NRGGAQNCATAGHNRHFDVG-VLPVTRDRWBOGVSPVKDQSCGS-WTFSTTGALRA 172
DB 121 SRGGAQNCATAGHNRHFDALPETYKDRDGVSPVKDQHGCGSCWPFSTTGLSEA 180
QY 173 AYTQLTGS--TLSEQQVLDCASAFNNFGC-GGLPSQAFYKYNKGIDTEQTYPYLVGNG 229
DB 181 RYTOATGPPVSLSEQQVLDCATRYNNFGCSGGLPSQAFYKYNKGIDTEQTYPYLVGNG 240
QY 230 ICNFKQENVGVKIDTSNITLGADELKHAAGLVPRVPSVAFVVKGNLYKGVVSSDTC 289
DB 241 ICHYKPNAGVKVLDSTNITLVADELKHAAGLVPRVPSVAFVVKGNLYKGVVSSDTC 300
QY 290 GRDPMVNHAVLAVGYGVEDGIPYWLKNSWGTNWDGNGYFMELGKMGCGVATCASYPI 349
DB 301 GTSMPDVNHAVLAVGYGVGVGPYWLKNSWGTNWDGNGYFMELGKMGCGVATCASYPI 360
QY 350 VA 351
DB 361 VA 362
RESULT 3
S71923
Cysteine proteinase (EC 3.4.22.-) - garden pea
C:Species: Pisum sativum (garden pea)
C>Date: 04-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 20-Jun-2000
C:Accession: S71923
R;Jones, C.G.; Tucker, G.A.; Lycett, G.W.
Biochim. Biophys. Acta 1296, 13-15, 1996
A;Title: Pattern of expression and characteristics of a cysteine proteinase cDNA from ge
A;Reference number: S71923; MUID:96350414; PMID:8765223
A;Accession: S71923
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-350 <JON>
A;Cross-references: EMBL:Z68291; NID:g1134881; PIDN:CAA92583.1; PID:g1134882

C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase
F;157,297,317/Active site: Cys, His, Asn #status predicted
Query Match 67.3%; Score 1254.5; DB 2; Length 350;
Best Local Similarity 67.2%; Pred. No. 2.7e-97;
Matches 240; Conservative 39; Mismatches 65; Indels 13; Gaps 6;
QY 1 MAAKLLFFLLFLVSALSVALAG--FREDNPIRSTQRPDS-IEPAILGVLSGRHAF 58
DB 1 MAQWSLLILVLCVAS---AAAGSFHSDNPIRMVS---DVEBQLQLQVIGESHAHVAF 53
QY 59 FARRYKSGSEIEIKRFGIFVENLAFIRSTRNRKDLSTYLGINQFADLTWEEFRTNRLG 118
DB 54 FARRYKSGSEIEIKRFGIFVENLAFIRSTRNRKDLSTYLGINQFADLTWEEFRTNRLG 113
QY 119 AAQNCSTAHGNHFRFVDVLPVTRDRWBOGVSPVKDQSCGS-WTFSTTGALRAAYTQL 177
DB 114 AAQNCSTAHGNHFRFVDVLPVTRDRWBOGVSPVKDQSCGS-WTFSTTGALRAAYTQL 173
QY 178 TGS--TLSEQQVLDCASAFNNFGC-GGLPSQAFYKYNKGIDTEQTYPYLVGNGICNFK 234
DB 174 FGKNSLSEQQVLDCASAFNNFGC-GGLPSQAFYKYNKGIDTEQTYPYLVGNGICNFK 233
QY 235 QENVGVKVIDSNITLGADELKHAAGLVPRVPSVAFVVKGNLYKGVVSSDTCGRDPM 294
DB 234 SEHVAVKVLGSVNITLGADELKHAAGLVPRVPSVAFVVKGNLYKGVVSSDTCGRDPM 293
QY 295 DVNHAVLAVGYGVEDGIPYWLKNSWGTNWDGNGYFMELGKMGCGVATCASYPIVA 351
DB 294 DVNHAVLAVGYGVEDGIPYWLKNSWGTNWDGNGYFMELGKMGCGVATCASYPIVA 350
RESULT 4
K3EH
Aleurain (EC 3.4.22.-) precursor - barley
C:Species: Hordeum vulgare (barley)
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 04-Oct-1996
C:Accession: A25492; B25492
R;Rogers, J.C.; Dean, D.; Heck, G.R.
Proc. Natl. Acad. Sci. U.S.A. 82, 6512-6516, 1985
A;Title: Aleurain: a barley thiol protease closely related to mammalian cathepsin H.
A;Reference number: A25492; MUID:86016732; PMID:3901004
A;Accession: A25492
A;Molecule type: DNA
A;Residues: 1-361 <ROG1>
A;Accession: B25492
A;Molecule type: mRNA
A;Residues: 1-361 <ROG2>
A;Cross-references: GB:X05167; GB:M11760; GB:M15209; GB:Y00107
C;Comment: Aleurain is synthesized by the aleurone cells stimulated by gibberellic or
C;Superfamily: papain
C;Keywords: aleurone cell; cysteine proteinase; germination; glycoprotein; hydrolase;
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-143/Domain: propeptide #status predicted <PRO>
F;144-361/Product: aleurain #status predicted <MAT>
F;165-208, 199-240, 298-348/Diulfide bonds: #status predicted
F;168, 307, 327/Active site: Cys, His, Asn #status predicted
F;188, 256/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 67.1%; Score 1252; DB 1; Length 361;
Best Local Similarity 67.3%; Pred. No. 4.6e-97;
Matches 241; Conservative 42; Mismatches 65; Indels 10; Gaps 7;
QY 3 AKLLFFLLFLVSALSVALA---GFEDNPIRSTQRPDS-IEPAILGVLSGRHAF 58
DB 4 ARVLLALAVLATAVAVASSSFDSPNPIRPTDRAASTLESVAVLGAHGRHAF 63
QY 59 FARRYKSGSEIEIKRFGIFVENLAFIRSTRNRKDLSTYLGINQFADLTWEEFRTNRLG 118
DB 64 FARRYKSGSEIEIKRFGIFVENLAFIRSTRNRKDLSTYLGINQFADLTWEEFRTNRLG 123
QY 119 AAQNCSTAHGNHFRFVDVLPVTRDRWBOGVSPVKDQSCGS-WTFSTTGALRAAYTQ 176

Db 124 AAQTCATLGNHMRDAALPETKDWREDDGIVSPKNOAHGCGSWFTSTTGALAAAYTQ 183
QY 177 LTGS--TLSEQQLVDCASAFNNFCG--GGLPSQAFYVYKNGGIDTEQTYPLGVMGICNF 233
Db 184 ATGNKISLSEQQLVDCAGFNFGNGGLPSQAFY--QYNGGIDTESYPKGVNGVCHY 242
QY 234 KOENVGVKVIDSINITLGADELKHAVALVLPVSVAFVYKNGGIDTEQTYPLGVMGICNF 293
Db 243 KAENAAVQLVDSVNITLNAEDLNKNAVLVLPVSVAFVYKNGGIDTESYPKGVNGVCHY 302
QY 294 MDVNHAVLAVGVGVEDGIPYMLIKNSGNTGNGYFPMELGKMGVATCASYPIVA 351
Db 303 DDVNHAVLAVGVGVEDGIPYMLIKNSGNTGNGYFPMELGKMGVATCASYPIVA 360

RESULT 5
S66348
cysteine proteinase (EC 3.4.22.-) senescence-associated precursor (clone SEN3) [similar]
C:Species: Lycopersicon esculentum (tomato)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S66348
R/Drake, R.; John, I.; Farrell, A.; Cooper, W.; Schuch, W.; Grierson, D.
Plant Mol. Biol. 30, 755-767, 1996
A>Title: Isolation and analysis of cDNAs encoding tomato cysteine proteases expressed during senescence
A:Reference number: S66348; MUID:96194458; PMID:8624407
A:Accession: S66348
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-356 <DRA>
A:Cross-references: GB:248736; NID:g1235544; PIDN:CAA88629.1; PID:g1235545
A:Superfamily: papain
C:Keywords: cysteine proteinase, hydrolase
F:163,303,323/Active site: Cys, His, Asn #status predicted

Query Match 67.0%; Score 1249; DB 2; Length 356;
Best Local Similarity 68.3%; Pred. No. 8.1e-97;
Matches 243; Conservative 35; Mismatches 70; Indels 8; Gaps 5;

QY 3 AKLLFFLLFVSALSVALAG---FEEDNPISVTPQDSTIEPAILGVLSGCRHAFHAF 59
Db 2 SRLSLVILVAGLFATALAGPATFADKNPIQVY--FPELENGILQVVGQTRLSALSF 60
QY 60 ARRYKSGSSEETKKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEPRTNRL 119
Db 61 AIRHKKRVDSEELKQREIFLDNLKMRSHNRKGLYKLGINEFTDLTWDEFKHLGA 120
QY 120 AONCSATAHGNHRFVDGVLVTRDWRQGVSPVVDQSGCS--WTFSTTGALAAAYTQ 178
Db 121 SONGSATTGKLNLTNVLPTKDWKRDGIVSPVKAQCGKCGSCWTFSTTGALAAAYTQ 180
QY 179 GS--TLSEQQLVDCASAFNNFCG--GGLPSQAFYVYKNGGIDTEQTYPLGVMGICNF 235
Db 181 KGKISLSEQQLVDCAGFNFGNGGLPSQAFYFVYKNGGIDTESYPKGVNGVCHY 240
QY 236 ENVGKVIDSINITLGADELKHAVALVLPVSVAFVYKNGGIDTESYPKGVNGVCHY 295
Db 241 ANIGKVIDSINITLGADELKHAVALVLPVSVAFVYKNGGIDTESYPKGVNGVCHY 300
QY 296 VNHAVLAVGVGVEDGIPYMLIKNSGNTGNGYFPMELGKMGVATCASYPIVA 351
Db 301 VNHAVLAVGVGVEDGIPYMLIKNSGNTGNGYFPMELGKMGVATCASYPIVA 356

RESULT 6
T47471
cysteine proteinase (EC 3.4.22.-) F18N11.70 [similarity] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 26-May-2000
C:Accession: T47471
R/Jordan, N.; Bangert, S.; Wiedemann, R.; Voss, H.; Unselid, M.; Mewes, H.W.; Lemcke, K.
submitted to the Protein Sequence Database, February 2000
A:Reference number: 224467

A:Accession: T47471
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-377 <DOR>
A:Cross-references: EMBL:AL132953
A:Experimental source: cultivar Columbia; BAC clone F18N11
C:Genetics:
A:Map position: 3
A:Introns: 65/2; 106/1; 144/3; 169/2; 257/3; 301/3; 345/2
A:Note: F18N11.70
C:Superfamily: papain
C:Keywords: cysteine proteinase, hydrolase
F:165,305,325/Active site: Cys, His, Asn #status predicted

Query Match 64.7%; Score 1207.5; DB 2; Length 377;
Best Local Similarity 65.9%; Pred. No. 2.6e-93;
Matches 228; Conservative 37; Mismatches 74; Indels 7; Gaps 4;

QY 1 MAAKLIF---FILFVSALSVALAGFEEDNPISVTPQDSTIEPAILGVLSGCRHAFHFA 57
Db 1 MSVKLNLSIIILLIFAAASKEIGFDESNPKMVSNDLHELEDTVVQILGOSRHVLSFS 60
QY 58 RFARRYKSGSSEETKKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEPRTNRL 117
Db 61 RTHRYGKGYQSVVEEMKLFPSVFKENLDLIRSTNKKGLSYKLSAQFADLTWEEPRYKL 120
QY 118 GAAQNCATAHGNHRFVDGVLVTRDWRQGVSPVVDQSGCS--WTFSTTGALAAAYTQ 176
Db 121 GAAQNCATLKGSHKITTEATVPTDKDWRDGVSPVKEQCHGCGSCWTFSTTGALAAAYH 180
QY 177 LTGS--TLSEQQLVDCASAFNNFCG--GGLPSQAFYVYKNGGIDTESYPKGVNGVCHY 233
Db 181 AFGKISLSEQQLVDCAGFNFGNGGLPSQAFYFVYKNGGIDTESYPKGVNGVCHY 240
QY 234 KOENVGVKVIDSINITLGADELKHAVALVLPVSVAFVYKNGGIDTESYPKGVNGVCHY 293
Db 241 SAKNIGVQVDSVNITLGADELKHAVALVLPVSVAFVYKNGGIDTESYPKGVNGVCHY 300
QY 294 MDVNHAVLAVGVGVEDGIPYMLIKNSGNTGNGYFPMELGKMGVATCASYPIVA 351
Db 301 MDVNHAVLAVGVGVEDGIPYMLIKNSGNTGNGYFPMELGKMGVATCASYPIVA 356

RESULT 7
KHRT
cathepsin H (EC 3.4.22.16) precursor - rat
N/Alternate names: cathepsin B3; cathepsin Ba
C:Species: Rattus norvegicus (Norway rat)
C>Date: 18-Apr-1984 #sequence_revision 30-Jun-1993 #text_change 21-Jul-2000
C:Accession: S00211; S05213; A00976; A60371
R/Ishidoh, K.; Imaizoh, S.; Emori, Y.; Ohno, S.; Kawasaki, H.; Kominami, E.;
FEBS Lett. 226, 33-37, 1987
A>Title: Molecular cloning and sequencing of cDNA for rat cathepsin H.
A:Reference number: S00211; MUID:88083632; PMID:3691815
A:Accession: S00211
A:Molecule type: mRNA
A:Residues: 1-333 <ISH>
A:Cross-references: GB:M36320; NID:g55885; PIDN:CAA68699.1; PID:g55886
R/Ishidoh, K.; Kominami, E.; Katunuma, N.; Suzuki, K.
FEBS Lett. 253, 103-107, 1989
A>Title: Gene structure of rat cathepsin H.
A:Reference number: S05213; MUID:89338722; PMID:2759235
A:Accession: S05213
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-286, 'A', 288-333 <ISH2>
R/Takio, K.; Towatari, T.; Katunuma, N.; Teller, D.C.; Titani, K.
Proc. Natl. Acad. Sci. U.S.A. 80, 3666-3670, 1983
A>Title: Homology of amino acid sequences of rat liver cathepsins B and H with that of
A:Reference number: A00976; MUID:83221657; PMID:6574504
A:Accession: A00976
A:Molecule type: protein
A:Residues: 114-333 <TAK>

R;Qian, F.; Frankfater, A.; Miller, R.V.; Chan, S.J.; Steiner, D.P.
 Int. J. Biochem. 22, 1457-1464, 1990
 A;Title: Molecular cloning of rat precursor cathepsin H and the expression of five lysos
 A;Reference number: A60371; MUID:91114797; PMID:2276418
 A;Accession: A60371
 A;Molecule type: mRNA
 A;Residues: 4-333 <Q1A>
 A;Cross-references: GB:M38135
 C;Comment: Cathepsin H is composed of a minichain and a large chain. The large chain may
 C;Genetics:
 A;Introns: 29/1; 39/3; 75/1; 98/3; 133/3; 162/3; 181/2; 208/3; 231/3; 267/2; 309/2
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; glycoprotein; hydrolase; lysosome
 F;1-21/Domain: signal sequence #status predicted <PRO>
 F;96-102/Product: cathepsin H minichain #status predicted <MCH>
 F;114-333/Product: cathepsin H large chain #status experimental <MCH>
 F;114-290/Product: cathepsin H heavy chain #status predicted <HCH>
 F;291-333/Product: cathepsin H light chain #status predicted <LCH>
 F;70-99/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;136-179, 170-212, 270-320/Disulfide bonds: #status predicted
 F;139, 279, 299/Active site: Cys, His, Asn #status predicted
 F;228/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 42.7%; Score 796; DB 1; Length 333;
 Best Local Similarity 52.1%; Pred. No. 6.2e-59;
 Matches 159; Conservative 50; Mismatches 88; Indels 8; Gaps 7;

QY 54 FHPARFARYKSGSBEIEKRFGEIVENLAFIRSTNEKDLSTYLGINQFADLTWEERF 113
 DB 31 FHETSMMKQKTYT-SREYSHRLQVFANNWRKIQIHNORNTFFKGLNPFSDMSFAEIK 89

QY 114 TNRL-GAAQNCATAGHNRHFDVGLVPTVRDREQ-IVSPVKDQSGCS-WTFSTTGAL 170
 DB 90 HKYLWSEPNQCSAT-KSNYLGRGTGPPSSMDWRKKGNVSPVKDQACGSCWTFSTTGAL 148

QY 171 EAYTQITGTS--TLSEQQLVDCASAFNNFGC-GGLPSQAFYKYNKGIDTQTPYLVG 227
 DB 149 ESAVAATAGSGWMTLASQQLVDCQNFNNHCQGLPSQAFYLYNKGIMGDSYPIYQK 208

QY 228 MGICNFQKQENVGVKVIDSINITLGADELKHAVLGRVPSVAFVVKGNLYKKGYSDD 287
 DB 209 NGQCKNPEKAVAFKVNVTITLDEAAVYEAVALYNPVSFAFVTEDFMYKSGYSSN 268

QY 288 TCGRDPMDVNHAVLAVGYGVEDGIPYWLKNSWGTWNGDGFQKELGKMGCVATCASY 347
 DB 269 SCHKTPDKVNHAVLAVGYGEQNGLLYWIYKNSWGSNGNNGYFLIERGKMGCLAACASY 328

QY 348 PIVAV 352
 DB 329 PIPQV 333

RESULT 8
 KHVH
 cathepsin H (EC 3.4.22.16) precursor [validated] - human
 C;Species: Homo sapiens (man)
 C;Date: 30-Sep-1990 #sequence revision 30-Jun-1991 #text change 08-Dec-2000
 C;Accession: S12486; S07634; S00818; S00322; A27011; C27011; S00635
 R;Fuchs, R.
 submitted to the EMBL Data Library, October 1989
 A;Reference number: S12486
 A;Accession: S12486
 A;Molecule type: mRNA
 A;Residues: 1-335 <FUC>
 A;Cross-references: EMBL:X16832; NID:929709; PIDN:CAA34734.1; PID:929710
 R;Fuchs, R.; Gassen, H.G.
 Nucleic Acids Res. 17, 9471, 1989
 A;Title: Nucleotide sequence of human preprocathepsin H, a lysosomal cysteine proteinase
 A;Reference number: S07634; MUID:90067944; PMID:2587265
 A;Accession: S07634
 A;Molecule type: mRNA
 A;Residues: 1-115 <F02>

A;Cross-references: EMBL:X16832
 R;Fuchs, R.; Machleidt, W.; Gassen, H.G.
 Biol. Chem. Hoppe-Seyler 369, 469-475, 1988
 A;Title: Molecular cloning and sequencing of a cDNA coding for mature human kidney cat
 A;Reference number: S00818; MUID:89076480; PMID:2849458
 A;Accession: S00818
 A;Molecule type: mRNA
 A;Residues: 88-335 <F03>
 A;Cross-references: EMBL:X07549; NID:929707; PIDN:CAA30428.1; PID:929708
 R;Ritonja, A.; Popovic, T.; Kocnik, M.; Machleidt, W.; Turk, V.
 FEBS Lett. 228, 341-345, 1988
 A;Title: Amino acid sequences of the human kidney cathepsins H and L.
 A;Reference number: S00322; MUID:88137635; PMID:3342889
 A;Accession: S00322
 A;Molecule type: protein
 A;Residues: 98-105; 114-178 'H', 180-305 'E', 307-335 <RIT>
 R;Machleidt, W.; Ritonja, A.; Popovic, T.; Kocnik, M.; Brzin, J.; Turk, V.; Machleidt,
 in Cysteine Proteinases and Their Inhibitors, Turk, V., ed., pp.3-18, Walter de Gruyter
 A;Title: Human cathepsins B, H and L: characterization by amino acid sequences and som
 A;Reference number: A27011
 A;Contents: annotation
 C;Comment: Cathepsin H is composed of a minichain and a large chain. The large chain m
 C;Genetics:
 A;Gene: GDB:CTSH
 A;Cross-references: GDB:120602; OMIM:116920
 A;Map position: 15q24-15q25
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; glycoprotein; hydrolase; lysosome
 F;1-22/Domain: signal sequence #status predicted <SIG>
 F;23-97/Domain: propeptide #status predicted <PRO>
 F;98-105/Product: cathepsin H minichain #status experimental <MCH>
 F;114-335/Product: cathepsin H large chain (minor form) #status experimental <MCH>
 F;116-335/Product: cathepsin H large chain #status experimental <MCH>
 F;116-232/Product: cathepsin H heavy chain #status experimental <HCH>
 F;293-335/Product: cathepsin H light chain #status experimental <LCH>
 F;72/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;101, 230/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F;102-327, 338-181, 172-214, 272-322/Disulfide bonds: #status predicted
 F;141, 281, 301/Active site: Cys, His, Asn #status predicted

Query Match 40.9%; Score 763; DB 1; Length 335;
 Best Local Similarity 52.1%; Pred. No. 3.6e-56;
 Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY 54 FHPARFARYKSGSBEIEKRFGEIVENLAFIRSTNEKDLSTYLGINQFADLTWEERF 113
 DB 33 FHFWSMKHKRTY-STEEVHRLQTFASNRKINAHNNHHTFKALNPFSDMSFAEIK 91

QY 114 TNRL-GAAQNCATAGHNRHFDVGLVPTVRDREQ-IVSPVKDQSGCS-WTFSTTGAL 170
 DB 92 HKYLWSEPNQCSAT-KSNYLGRGTGPPSSMDWRKKGNVSPVKDQACGSCWTFSTTGAL 150

QY 171 EAYTQITGTS--EQQLVDCASAFNNFGC-GGLPSQAFYKYNKGIDTQTPYLVG 227
 DB 151 ESAIATATGRMLSLAEQQLVDCQDFNNYQCGGLPSQAFYLYNKGIMGDSYPIYQK 210

QY 228 MGICNFQKQENVGVKVIDSINITLGADELKHAVLGRVPSVAFVVKGNLYKKGYSDD 287
 DB 211 DGYCKFGPKAIGFVKDANITIVDEAMVEAVALYNPVSFAFVTEDFMYRTGIYSST 270

QY 288 TCGRDPMDVNHAVLAVGYGVEDGIPYWLKNSWGTWNGDGFQKELGKMGCVATCASY 347
 DB 271 SCHKTPDKVNHAVLAVGYGEKNGIPYIVKNSWGPQWGMNGYFLIERGKMGCLAACASY 330

QY 348 PIVAV 352
 DB 331 PIPLV 335

RESULT 9
 S43991
 cathepsin L-like proteinases (EC 3.4.22.-) - liver fluke
 C;Species: Fasciola hepatica (liver fluke)

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 04-Feb-2000
 C;Accession: S43991
 R;Wijffels, G.L.; Panaccio, M.; Salvatore, L.; Wilson, L.; Walker, I.D.; Spithill, T.W. Biochem. J. 295, 781-790, 1994
 A;Title: The secreted cathepsin L-like proteinases of the trematode, Fasciola hepatica, A;Reference number: S43991; MUID:94250242; PMID:8192668
 A;Accession: S43991
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-326 <WJ>
 A;Cross-references: GB:L33771; NID:G497699; PIDN:AAA29136.1; PID:G497700
 A;Note: the sequence from Fig. 6 is inconsistent with that from Fig. 5 in having 57-Pro
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; hydrolase
 F;132,269,289/Active site: Cys, His, Asn #status predicted

Query Match 35.2%; Score 656.5; DB 2; Length 326;
 Best Local Similarity 44.6%; Pred. No. 2.8e-47;
 Matches 144; Conservative 52; Mismatches 102; Indels 25; Gaps 13;

Qy 44 LGVLGSCRHAFHFAFARRYKSYGSEEEIKKRFIFVENLAFIRSTN-RKDL---SYTL 99
 Db 11 VGVLSNDLWH--QWKMYNKYNGADDQHR-NINEKNVKHQEHNLRLHDLGLVYTL 67

Qy 100 GINQFADLTWEFFRTNRLGAQNC-ATAHG-----NHRFVGVLPVTRDWRQGIYSPV 153
 Db 68 GLNQFTDWTFFEFKAKYLTMSRASDILSHGVPEYANRAV----PKIDWRSYGYTEV 123

Qy 154 KDQSCGS-WTFSITGTALEAY--TQLTGSTLSQQLVDCASAFNFGC-GGLPSQAFY 209
 Db 124 KDQNGCSGSAFSTITGTEGQYMKNRTSISFSQQLVDCSGPGNGGCGGLMENAYQY 183

Qy 210 VKYNGGIDTEQTYPYLGVMGICNFKQENVGVKVIDSINITLGADELKHAVALVVRPVVA 269
 Db 184 LK-QFGLTESSTPYTAVEGQCRYNKQGVAKVTGYTYVHSGSEVELKLVGARRPAVA 242

Qy 270 FEVKGENLYKGVYSSDTQGRDPMVNHAVLAVGYVEDGIPYWLKNSWGTNGWNGY 329
 Db 243 VDVSDFPMYRSGIYQSGTC--SPLRVNHAVLAVGYTGQGTDTWIVNSWGTWYGERGY 300

Qy 330 FKMLGK-NMCGVATCASYPTVA 351
 Db 301 IRVARNRGNMCGIASLASLPVA 323

RESULT 10
 T03941
 cysteine proteinase (EC 3.4.22.-) precursor - common tobacco
 C;Species: Nicotiana tabacum (common tobacco)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 C;Accession: T03941
 R;Becker, C.; Muentz, K.
 submitted to the EMBL Data Library, September 1997
 A;Description: CDNA cloning of a CPRI-homologous proteinase from germinating tobacco seed
 A;Reference number: Z15148
 A;Accession: T03941
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-374 <BC>
 A;Cross-references: EMBL:Z99173; PIDN:CAB16317.1
 A;Experimental source: clone TCPR1
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; hydrolase
 F;1-20/Domain: signal sequence #status predicted <SIG>
 F;21-374/Product: cysteine proteinase #status predicted <MAT>
 F;164,299,319/Active site: Cys, His, Asn #status predicted

Query Match 33.0%; Score 616; DB 2; Length 374;
 Best Local Similarity 40.4%; Pred. No. 8.3e-44;
 Matches 153; Conservative 58; Mismatches 112; Indels 56; Gaps 17;

Qy 1 MAKLIFLLF-LVSALSA---LAGFEENPIRSTVTRPDSTIEPAILGVLSCHAFH 55

Db 1 MAKTIITLLFALPSSLSYAIMDSIIDYKNNHYARKWTLQSD--EDQVKN-----R 49

Qy 56 FARFARRYKSYGSEEEIKKRFIFVENLAFIRSTNRK-DLSYTLGINQFADLTWEERT 114
 Db 50 YEWLAHGRAYNALGEKEKREIFKDLNRFIEGHNSGNRYKYVGLNQFADLTWEERT 109

Qy 115 NELGAAQNCASATAGNHRPV-----DGVLPVTRDWRQGIYSPVVKDQSGCS 161
 Db 110 MYLGTKSD-----ARRRFVSKNPSQRYASRPNELMPHSVDWRKGVAPVKNQSGCS 163

Qy 162 -WTFSTTGTALEAAYTQLTGS--TLSEQQLVDCASAFNFGC-GGLPSQAFYVKNYGGID 217
 Db 164 CWFSTVAAVEGINQIVGEMITLSEQLVDC-DRVQNSGCGGLMDYAFVFIISNGND 222

Qy 218 TEQTYPYLGVMGICNFKQENVGVKVIDSINITLGADELKHAVALVVRPVSAFVYV-KGF 276
 Db 223 TEKHPYRGVEGRCDPVKRYKVSIDGYEDVPRNERALQKAVA-HQPVCAVAEASGRAF 281

Qy 277 NLYKGVYSSDTQGRDPMVNHAVLAVGYVEDGIPYWLKNSWGTNGWNGYFKWE--- 333
 Db 282 QLYSGVFTGE-CGE---EVDHGVVYVVGSGEDGVDYIVRNSWGTNGWNGYVKNERNV 337

Qy 334 ----LGNMCGVATCASYP 348
 Db 338 KKSHLGK--CGIMTEASYP 354

RESULT 11
 JCS443
 cathepsin L-like cysteine proteinase (EC 3.4.22.-) cl [similarity] - Maize weevil
 C;Species: Sitophilus zeamais (maize weevil)
 C;Date: 17-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000
 C;Accession: JCS443
 R;Matsumoto, I.; Emori, Y.; Abe, K.; Arai, S.
 J. Biochem. 121, 464-476, 1997
 A;Title: Characterization of a gene family encoding cysteine proteinases of Sitophilus
 nd germ cells.
 A;Reference number: JCS441; MUID:97279038; PMID:9133615
 A;Accession: JCS443
 A;Molecule type: mRNA
 A;Residues: 1-338 <MAT>
 A;Cross-references: GB:D82884; NID:G2804261; PIDN:BAA24442.1; PID:G2804262
 C;Comment: This enzyme acts as a secreted or lysosomal proteinase. It has various funct
 y organs, and cleaning in the malpighian tubule.
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; hydrolase
 F;145,284,305/Active site: Cys, His, Asn #status predicted

Query Match 32.9%; Score 613.5; DB 2; Length 338;
 Best Local Similarity 43.6%; Pred. No. 1.2e-43;
 Matches 137; Conservative 49; Mismatches 107; Indels 21; Gaps 10;

Qy 56 FARFARRYKSYGSEEEIKKRFIFVENLAFIRSTNR-----KDLSTYTLGINQFADLTWE 111
 Db 27 WSPFMQHSKNYDSTETEFPRMKIIPWENAHKVAKNKLPSCQFVKFGLNAYADMLHHE 86

Qy 112 FRTNLGAQNCASATAGNH-----RFVDGV---LPVTRDWRQGIYSPVVKDQSGCS-W 162
 Db 87 FVSTLNGFNKTNILKGSGLNDLDAVRFISPANVKLPDVTVDWRDKGAVTEVDKQCHGSCW 146

Qy 163 TFSITGTALEAAYTQLTGS--TLSEQQLVDCASAFNFGC-GGLPSQAFYVKNYGGIDTE 219
 Db 147 SFSATSLGEGQHFRTKGLVLSLEQNLVDCSGRYGNGGCGGLMDNAYFIYKDNIGDITE 206

Qy 220 QTYPYLGVNMGICNFKQENVGVKVIDSINITLGADELKHAVALVVRPVSAFVYV-KGFNL 278
 Db 207 KSVPYLADEKCHYKAQNSGATDKGFVDIEANEDDLKAAVATVGPVSVIAIDASHETFOL 266

Qy 279 YKGVYSSDTQGRDPMVNHAVLAVGYV-EDGIPYWLKNSWGTNGWNGYFKWELGK- 336
 Db 267 YSDGVYSDPECSQBLD--HGVLVVGYGTSDDQDYWLKNSWGSGLNGYIKWARKQD 324

Qy 337 NMCGVATCASYPIV 350

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 21:10:38 ; Search time 25 Seconds
(without alignments)
733.148 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAKKLFFLLVLSLSVAL.....ELGRKMGVATCASYPIVAV 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1322.5	70.9	360	1	Q10717 zea mays (m
2	1281.5	68.7	362	1	P25778 oryza sativ
3	1274.5	68.3	362	1	P05167 hordeum vul
4	1249	67.0	356	1	Q40143 lycopersico
5	796	42.7	333	1	P00786 rattus norv
6	792	42.5	335	1	O46427 sus scrofa
7	765	41.0	333	1	P49935 mus musculu
8	759	40.7	335	1	P09668 homo sapien
9	609	32.7	462	1	P43297 arabidopsis
10	603.5	32.4	339	1	Q28636 sarcophaga
11	594.5	31.9	323	1	P25782 homarus ame
12	589	31.6	344	1	P54640 dictyosteli
13	587	31.5	471	1	P25777 oryza sativ
14	586	31.4	322	1	P13277 homarus ame
15	581	31.2	341	1	Q95029 drosophila
16	570	30.6	328	1	P25251 brassica na
17	563.5	30.2	363	1	P25804 pisum sativ
18	558	29.9	376	1	P04989 dictyosteli
19	556.5	29.8	321	1	P25784 homarus ame
20	556	29.8	380	1	P00785 actinidia c
21	554.5	29.7	458	1	P25776 oryza sativ
22	550	29.5	364	1	Q9aut0 arabidopsis
23	543.5	29.1	362	1	Q9it77 arabidopsis
24	541.5	29.0	334	1	O60911 homo sapien
25	539.5	28.9	371	1	Q9seu9 arabidopsis
26	539	28.9	361	1	P43295 arabidopsis
27	536.5	28.8	368	1	P43296 arabidopsis
28	531.5	28.5	329	1	O77641 macaca fasc
29	528.5	28.3	329	1	P43236 oryctolagus
30	528.5	28.3	333	1	Q9i014 mus musculu
31	527.5	28.3	329	1	P43235 homo sapien
32	527.5	28.3	371	1	Q10716 zea mays (m
33	527	28.3	362	1	P12412 vigna mungo

ALIGNMENTS

RESULT 1

CYS2_MAIZE STANDARD; PRT; 360 AA.

AC Q10717; 1 CATK_RAT

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cysteine proteinase 2 precursor (EC 3.4.22.-).

GN CCP2.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC PACAD clade; Panicoideae; Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=96004895; PubMed=7548211;

RA Domoto C., Watanabe H., Abe M., Abe K., Arai S.;

RT "Isolation and characterization of two distinct cDNA clones encoding

corn seed cysteine proteinases.";

RL Biochim. Biophys. Acta 1263:241-244(1995).

CC -!- FUNCTION: Involved in the degradation of the storage protein zein.

CC -!- SUBCELLULAR LOCATION: Vacuolar (Probable).

CC -!- TISSUE SPECIFICITY: Expressed at the onset of germination.

CC -!- SIMILARITY: TO OTHER EUKARYOTIC THIOL PROTEASES.

CC -!- SIMILARITY: Belongs to peptidase family C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D45403; BAA08245.1; -.

DR EMBL; S59598; S59598.

DR HSP; O46427; 8FCH.

DR MEROPS; C01.041; -.

DR MaizeDB; 25431; -.

DR InterPro; IPR000169; SHprot acsite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00845; Pept_C1; 1.

DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.

DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.

DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.

KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal.

FT SIGNAL 1 19 POTENTIAL.

FT PROPEP 20 142 ACTIVATION PEPTIDE (POTENTIAL).

FT CHAIN 143 360 CYSTEINE PROTEINASE 2.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 167 167 BY SIMILARITY.

FT ACT_SITE 307 307 BY SIMILARITY.
 FT ACT_SITE 327 327 BY SIMILARITY.
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 360 AA; 39199 MW; 8B5788F1B0C3FC1C CRC64;

Query Match 70.9%; Score 1322.5; DB 1; Length 360;
 Best Local Similarity 71.7%; Pred. No. 7.4e-98;
 Matches 258; Conservative 28; Mismatches 65; Indels 9; Gaps 6;

QY 1 MAAKLFFLLFVLSALVAL--AGFEEDNPIRSVTQRPDS-IEPAILGLVGSCHAF 57
 DB 1 MVRRLFLVAVVLDATAVAVNSGFADSNFIRPVTDRASALESVFVAALGRTRDALFA 60

QY 58 RPARRYKGVSEEEIKKRGIFVENLAFTSTNRKDLSTYLTGINQFADLTWEERTNRL 117
 DB 61 RFARVYKGVSEAAEVHKKRIFSESLQVLRNKGSLYRGLINRKFADMSWEEFQA 120

QY 118 GAAQNCATAGNHRP--VDGVLPTVRDWRQGVSPVKDQSCGS-WTFSTTGALAA 174
 DB 121 GAAQNCATLTGNHRMRAAVALPETKQWREDGIVSPVKQGHGCSWTFSTTGALAA 180

QY 175 TQLTGS--TLSEQQLVDCASAFNFGC--GGLPSQAFYKYNKGIDTETQTPYLGVMGIC 231
 DB 181 TQATGRPISLSEQQLVDCGFAPNFGCNGGLPSQAFYKYNKGIDTETQTPYLGVMGIC 240

QY 232 NFKQENVGVKVIDSINITLGADELKHAVALGVRPVSVAPEVVKGNLKYGVYSDDTCGR 291
 DB 241 KFNENVGVKVLDSVNITLGADELKHAVALGVRPVSVAPEVITGRFYKSGYTTSDHC 300

QY 292 DPMVDNVHVLAVGYGVEDGIPWLIKNSWGTWNGDGYFKMELGKMGCVATCASYPIVA 351
 DB 301 TMDVDNVHVLAVGYGVEDGVPYWLKNSWGADWDGDEGYFKMGMKMGCVATCASYPIVA 360

RESULT 2

ORIC-ORYSA
 ID -ORYC-ORYSA STANDARD; PRT; 362 AA.

AC P25778;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Oryza gamma chain precursor (EC 3.4.22.-).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare; TISSUE=Seed;
 RX MEDLINE=91358494; PubMed=1885617;
 RA Watanabe H., Abe K., Emori Y., Hosoyama H., Arai S.;
 RT "Molecular cloning and gibberellin-induced expression of multiple
 RT cysteine proteinases of rice seeds (oryzains).";
 RL J. Biol. Chem. 266:16897-16902(1991).
 CC -1- TISSUE SPECIFICITY. Expressed only in seeds.
 CC -1- INDUCTION: By gibberellic acid (GA).
 CC -1- SIMILARITY: Belongs to peptidase family C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; D90408; BAA14404.1; --
 CC PIR; J00390; KHRZOG.
 CC HGSP; P07711; 1CUL.
 CC MEROB; C01.041; --
 CC Gramene; P25778; --

InterPro; IPR000669; Peptidase_C1.
 InterPro; IPR000169; SHprot_acsite.
 Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIA.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal.

FT SIGNAL 1 24 POTENTIAL..
 FT PROPEP 25 144 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 145 362 ORYZAIN GAMMA CHAIN.
 FT ACT_SITE 169 169 BY SIMILARITY.
 FT ACT_SITE 309 309 BY SIMILARITY.
 FT ACT_SITE 329 329 BY SIMILARITY.
 FT DISULFID 166 209 BY SIMILARITY.
 FT DISULFID 200 242 BY SIMILARITY.
 FT DISULFID 300 350 BY SIMILARITY.
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 362 AA; 39204 MW; 73E573EB1DA10B9D CRC64;

Query Match 68.7%; Score 1281.5; DB 1; Length 362;
 Best Local Similarity 66.9%; Pred. No. 1.3e-94;
 Matches 242; Conservative 45; Mismatches 64; Indels 11; Gaps 6;

QY 1 MAAKLFFLLF-----VSALSVALAGFEEDNPIRSVTQRPDS-IEPAILGLVGSCHAF 54
 DB 1 MAHRRITLLLAVAATAVSAVAASGDFDSDNPIRSVTDHAAALESTVIAALGRTRGAL 60

QY 55 HFARFARYKGVSEEEIKKRGIFVENLAFTSTNRKDLSTYLTGINQFADLTWEERTN 114
 DB 61 RFARFARVHKGKRGDAEAVQRRIFSESLVLRNRRGLPYRLGLNRRFADMSWEEFQA 120

QY 115 NRLGAAQNCATAGNHRPVDG-VLPVTRDWRQGVSPVKDQSCGS-WTFSTTGALAA 172
 DB 121 SRLGAAQNCATLTAGNHRMEDAPALPETKQWREDGIVSPVKDQGHGCSWTFSTTGALAA 180

QY 173 AYTQLTGS--TLSEQQLVDCASAFNFGC--GGLPSQAFYKYNKGIDTETQTPYLGVMG 229
 DB 181 RYTOATGPVPSLSEQQLADCATRYNFGCGGLPSQAFYKYNKGIDTETQTPYLGVMG 240

QY 230 ICNFKQENVGVKVIDSINITLGADELKHAVALGVRPVSVAPEVVKGNLKYGVYSDDTC 289
 DB 241 ICHTKPNAGVKVLDSVNITLGADELKHAVALGVRPVSVAPEVVKGNLKYGVYSDDTC 300

QY 290 GRDPMVDNVHVLAVGYGVEDGIPWLIKNSWGTWNGDGYFKMELGKMGCVATCASYPI 349
 DB 301 GTSMDNVHVLAVGYGVENGVPYWLKNSWGADWDGDEGYFKMGMKMGCVATCASYPI 360

QY 350 VA 351
 DB 361 VA 362

RESULT 3
 ALEU HORVU
 ID ALEU HORVU STANDARD; PRT; 362 AA.

AC P05167;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Thiol protease aleurain precursor (EC 3.4.22.16).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86016732; PubMed=3901004;
 RA Rogers J.C., Dean D., Heck G.R.;

"Aleurain: a barley thiol protease closely related to mammalian cathepsin H.",
Proc. Natl. Acad. Sci. U.S.A. 82:6512-6516 (1985).
[2]

RP REVISIONS.

RA Rogers J.C.;
Submitted (MAR-1987) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins, acting as an
aminopeptidase (notably, cleaving Arg-[Xaa bonds] as well as an
endopeptidase.

CC -1- SUBCELLULAR LOCATION: VACUOLE-LIKE SUBCELLULAR COMPARTMENT.

CC -1- INDUCTION: Aleurain is synthesized by the aleurone cells

CC stimulated by gibberellic or abscisic acid.

CC -1- SIMILARITY: Belongs to peptidase family C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
send an email to license@isb-sib.ch).

CC EMBL; X05167; CAA28804.1; -.

DR HSSP; O46427; 8PCH.

DR MEROPS; C01.041; -.

DR InterPro; IPR000668; Peptidase C1.

DR InterPro; IPR000169; SHprot acSite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept C1; 1.

DR PROSITE; PS00139; THIOI PROTEASE CYS; 1.

DR PROSITE; PS00639; THIOI PROTEASE_HIS; 1.

DR PROSITE; PS00640; THIOI PROTEASE_ASN; 1.

KW Hydrolase; Thiol protease; Glycoprotein; Zymogen; Signal; Germination.

FT SIGNAL 1 22

FT PROPEP 23 143 ACTIVATION PEPTIDE (POTENTIAL).

FT CHAIN 144 362 THIOI PROTEASE ALEURAIN.

FT ACT_SITE 168 168 BY SIMILARITY.

FT ACT_SITE 308 308 BY SIMILARITY.

FT ACT_SITE 328 328 BY SIMILARITY.

FT DISULFID 165 208 BY SIMILARITY.

FT DISULFID 199 241 BY SIMILARITY.

FT DISULFID 299 349 BY SIMILARITY.

FT CARBOHYD 188 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

QY 294 MDVNHAVLAVGVGVEDGIPYWLKNSWGTNWDNGYFKMELGKMGVATCASYPIVA 351
DB 304 DDVNEHVLAVGVGVGVEDGIPYWLKNSWGTNWDNGYFKMELGKMGVATCASYPIVA 361

RESULT 4

CYS3_LYCES

ID CYS3_LYCES STANDARD; PRT; 356 AA.

AC Q40143;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cysteine proteinase 3 precursor (EC 3.4.22.-).

GN CYP-3.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC lamids; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4081;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Ailsa Craig; TISSUE=Leaf;

RC MEDLINE=96194458; PubMed=8244407;

RA Drake R., John I., Farrell A., Cooper W., Schuch W., Grierson D.;

RT "Isolation and analysis of cDNAs encoding tomato cysteine proteases

RT expressed during leaf senescence.";

RL Plant Mol. Biol. 30:755-767(1996).

CC -1- SUBCELLULAR LOCATION: Vacuolar (Probable).

CC -1- TISSUE SPECIFICITY: Predominantly expressed in stem and root.

CC -1- SIMILARITY: Belongs to peptidase family C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
send an email to license@isb-sib.ch).

CC EMBL; Z48736; CAA88629.1; -.

DR PIR; S66348; S66348.

DR HSSP; O46427; 8PCH.

DR MEROPS; C01.041; -.

DR InterPro; IPR000668; Peptidase C1.

DR InterPro; IPR000169; SHprot acSite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept C1; 1.

DR PROSITE; PS00139; THIOI PROTEASE CYS; 1.

DR PROSITE; PS00639; THIOI PROTEASE_HIS; 1.

DR PROSITE; PS00640; THIOI PROTEASE_ASN; 1.

KW Hydrolase; Thiol protease; Glycoprotein; Zymogen; Signal.

FT SIGNAL 1 16

FT PROPEP 17 138 ACTIVATION PEPTIDE (POTENTIAL).

FT CHAIN 139 356 CYSINE PROTEINASE 3.

FT ACT_SITE 163 163 BY SIMILARITY.

FT ACT_SITE 303 303 BY SIMILARITY.

FT ACT_SITE 323 323 BY SIMILARITY.

FT DISULFID 160 203 BY SIMILARITY.

FT DISULFID 194 236 BY SIMILARITY.

FT DISULFID 294 349 BY SIMILARITY.

FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 67.0%; Score 1249; DB 1; Length 356;

Best Local Similarity 68.3%; Pred. No. 5.1e-92;

Matches 243; Conservative 35; Mismatches 70; Indels 8; Gaps 5;

QY 3 AKLLFLLFLVLSALVALAG---FEEDNPIRSVTQRPDSIEPAIILGVLSGCRHAFHARF 59

DB 2 SRLSLVLVLVAGLAFATAGPATFADKNPIROV-FPDELENGILQVGVQTRSAISAF 60

RESULT 6

CATH_PIG STANDARD; PRT; 335 AA.

AC O46427; 121

DT 15-JUL-1999 (Rel. 38, Created) 122

DT 15-JUL-1999 (Rel. 38, Last sequence update) 123

DT 10-OCT-2003 (Rel. 42, Last annotation update) 124

DE Cathepsin H precursor (EC 3.4.22.16). 125

GN CTSH. 126

OS Sus scrofa (Pig). 127

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; 128

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus. 129

OX NCBI_TaxID=9823; 130

RN [1] 131

RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS). 132

RC TISSUE=Peripheral blood, and Spleen; 133

RX MEDLINE=98154318; PubMed=9493267; 134

RA Guncar G., Podobnik M., Pungercar J., Strukelj B., Turk V., Turk D.; 135

RT "Crystal structure of porcine cathepsin H determined at 2.1-A 136

RT resolution; location of the mini-chain C-terminal carboxyl group 137

RT defines cathepsin H aminopeptidase function."; 138

RL Structure 6:51-61(1998). 139

CC -!- FUNCTION: Important for the overall degradation of proteins in 140

CC lysosomes. 141

CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins, acting as an 142

CC aminopeptidase (notably, cleaving Arg-Xaa bonds) as well as an 143

CC endopeptidase. 144

CC -!- SUBUNIT: Composed of a minichain and a large chain. The large 145

CC chain may be split into heavy and light chain. All chains are held 146

CC together by disulfide bonds. 147

CC -!- SUBCELLULAR LOCATION: Lysosomal. 148

CC -!- SIMILARITY: Belongs to peptidase family C1. 149

CC ----- 150

CC This SWISS-PROT entry is copyright. It is produced through a collaboration 151

CC between the Swiss Institute of Bioinformatics and the EMBL outstation - 152

CC the European Bioinformatics Institute. There are no restrictions on its 153

CC use by non-profit institutions as long as its content is in no way 154

CC modified and this statement is not removed. Usage by and for commercial 155

CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/> 156

CC or send an email to license@ebi.ac.uk). 157

CC ----- 158

DR EMBL; AF001169; AAB93957.1; -. 159

DR PDB; 8FCF; 13-JAN-99. 160

DR PDB; 1NB3; 18-FEB-03. 161

DR MEROPS; C01.040; -. 162

DR InterPro; IPR000568; Peptidase C1. 163

DR InterPro; IPR000169; SHprot_acsite. 164

DR Pfam; PF00112; Peptidase_C1; 1. 165

DR PRINTS; PR00705; PAPAIN. 166

DR ProDom; PD000158; Peptidase C1; 1. 167

DR SMART; SM00645; Pept C1; 1. 168

DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1. 169

DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1. 170

DR PROSITE; PS00840; THIOL_PROTEASE_ASN; 1. 171

KW Hydrolase; Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal; 172

KW 3D-structure. 173

FT SIGNAL 1 22 POTENTIAL. 174

FT PROPEP 23 97 POTENTIAL. 175

FT CHAIN 98 105 CATHEPSIN H MINICHAIN. 176

FT PROPEP 107 115 177

FT CHAIN 116 135 CATHEPSIN H. 178

FT ACT_SITE 141 141 179

FT ACT_SITE 281 281 180

FT ACT_SITE 301 301 181

FT DISULFID 102 327 182

FT DISULFID 138 181 183

FT DISULFID 172 214 184

FT DISULFID 272 322 185

FT CARBOHYD 72 72 186

FT CARBOHYD 101 101 187

FT CARBOHYD 230 230 188

FT N-LINKED (GLCNAC. . .) (POTENTIAL). 189

FT N-LINKED (GLCNAC. . .) (POTENTIAL). 190

FT N-LINKED (GLCNAC. . .) (POTENTIAL). 191

RESULT 7

CATH_MOUSE STANDARD; PRT; 333 AA.

ID_CATH_MOUSE STANDARD; PRT; 333 AA.

AC P49935; 333

DT 01-OCT-1996 (Rel. 34, Created) 335

FT STRAND 120

FT HELIX 121

FT STRAND 122

FT STRAND 123

FT TURN 124

FT TURN 125

FT HELIX 126

FT STRAND 127

FT STRAND 128

FT HELIX 129

FT TURN 130

FT TURN 131

FT STRAND 132

FT STRAND 133

FT TURN 134

FT TURN 135

FT STRAND 136

FT STRAND 137

FT TURN 138

FT TURN 139

FT STRAND 140

FT STRAND 141

FT TURN 142

FT TURN 143

FT STRAND 144

FT STRAND 145

FT TURN 146

FT TURN 147

FT STRAND 148

FT STRAND 149

FT TURN 150

FT TURN 151

FT STRAND 152

FT STRAND 153

FT TURN 154

FT TURN 155

FT STRAND 156

FT STRAND 157

FT TURN 158

FT TURN 159

FT STRAND 160

FT STRAND 161

FT TURN 162

FT TURN 163

FT STRAND 164

FT STRAND 165

FT TURN 166

FT TURN 167

FT STRAND 168

FT STRAND 169

FT TURN 170

FT TURN 171

FT STRAND 172

FT STRAND 173

FT TURN 174

FT TURN 175

FT STRAND 176

FT STRAND 177

FT TURN 178

FT TURN 179

FT STRAND 180

FT STRAND 181

FT TURN 182

FT TURN 183

FT STRAND 184

FT STRAND 185

FT TURN 186

FT TURN 187

FT STRAND 188

FT STRAND 189

FT TURN 190

FT TURN 191

FT STRAND 192

FT STRAND 193

FT TURN 194

FT TURN 195

FT STRAND 196

FT STRAND 197

FT TURN 198

FT TURN 199

FT STRAND 200

FT STRAND 201

FT TURN 202

FT TURN 203

FT STRAND 204

FT STRAND 205

FT TURN 206

FT TURN 207

FT STRAND 208

FT STRAND 209

FT TURN 210

FT TURN 211

FT STRAND 212

FT STRAND 213

FT TURN 214

FT TURN 215

FT STRAND 216

FT STRAND 217

FT TURN 218

FT TURN 219

FT STRAND 220

FT STRAND 221

FT TURN 222

FT TURN 223

FT STRAND 224

FT STRAND 225

FT TURN 226

FT TURN 227

FT STRAND 228

FT STRAND 229

FT TURN 230

FT TURN 231

FT STRAND 232

FT STRAND 233

FT TURN 234

FT TURN 235

FT STRAND 236

FT STRAND 237

FT TURN 238

FT TURN 239

FT STRAND 240

FT STRAND 241

FT TURN 242

FT TURN 243

FT STRAND 244

FT STRAND 245

FT TURN 246

FT TURN 247

FT STRAND 248

FT STRAND 249

FT TURN 250

FT TURN 251

FT STRAND 252

FT STRAND 253

FT TURN 254

FT TURN 255

FT STRAND 256

FT STRAND 257

FT TURN 258

FT TURN 259

FT STRAND 260

FT STRAND 261

FT TURN 262

FT TURN 263

FT STRAND 264

FT STRAND 265

FT TURN 266

FT TURN 267

FT STRAND 268

FT STRAND 269

FT TURN 270

FT TURN 271

FT STRAND 272

FT STRAND 273

FT TURN 274

FT TURN 275

FT STRAND 276

FT STRAND 277

FT TURN 278

FT TURN 279

FT STRAND 280

FT STRAND 281

FT TURN 282

FT TURN 283

FT STRAND 284

FT STRAND 285

FT TURN 286

FT TURN 287

FT STRAND 288

FT STRAND 289

FT TURN 290

FT TURN 291

FT STRAND 292

FT STRAND 293

FT TURN 294

FT TURN 295

FT STRAND 296

FT STRAND 297

FT TURN 298

FT TURN 299

FT STRAND 300

FT STRAND 301

FT TURN 302

FT TURN 303

FT STRAND 304

FT STRAND 305

FT TURN 306

FT TURN 307

FT STRAND 308

FT STRAND 309

FT TURN 310

FT TURN 311

FT STRAND 312

FT STRAND 313

FT TURN 314

FT TURN 315

FT STRAND 316

FT STRAND 317

FT TURN 318

FT TURN 319

FT STRAND 320

FT STRAND 321

FT TURN 322

FT TURN 323

FT STRAND 324

FT STRAND 325

FT TURN 326

FT TURN 327

FT STRAND 328

FT STRAND 329

FT TURN 330

FT TURN 331

FT STRAND 332

FT STRAND 333

SQ SEQUENCE 335 AA; 37454 MW; F728E845292C3B55 CRC64;

Query Match 42.5%; Score 792; DB 1; Length 335;

Best Local Similarity 54.3%; Pred. No. 9.9e-56;

Matches 165; Conservative 37; Mismatches 94; Indels 8; Gaps 7;

QY 55 HFARFARYKSYGSSEEEIKKRGIFVEMLAIFRSTNRKDLSTYLGINOFADLTWEEFRT 114

Db 34 HFKSMVMVQKKY-SLEEVYHRLQVPSNWRKINAHNAGNHTPKLGLNQFSDMSFDEIRH 92

QY 115 NRI-GAAQCSATAHGNHRFVGVLTVDREOG-IVSPVKDQSGCS-WTFSTTTGALE 171

Db 93 KYLWSEFQNCAT-KGNYLRGTGTPYPPSMDWRKKNFVSPVKQSGCGSCWTFSTTTGALE 151

QY 172 AAYTQLTGSTLS--EQQLVDCASAFNNFC-GGLPSQAFYKYKNGGIDTEQTYPYLGVN 228

Db 152 SAVAIATGKMLSLAEQQLVDCAQNFNNHGGQGLPSQAFYIRYNKNGIMGEDTYPYKQD 211

QY 229 GICNFQENGVGVKVIDSINITLGADELKHAHGLVPSVAPEVVKGNLYKKGVYSSDT 288

Db 212 DHCKFPDKAIAFAFKDVANITMDEEAMVEALVNPVSFAFEVTDNDFLMIRKGIYSSTIS 271

QY 289 CGRDPMDVNAHVLAVGVGVDEGIPYWLIIKNSWGTNMGNDGYPFMELGKMGCVATCASYP 348

Db 272 CHKTPDKVNAHVLAVGVGEENGIPYWIKNKSWGPMQGMNGYFLIERGKMGCLAACASYP 331

QY 349 IVAV 352

Db 332 IPLV 335

FT	CARBOHYD	228	228	N-LINKED (GLCNAC...)	(POTENTIAL).
FT	CONFLICT	137	137	A -> G (IN REF.2).	
FT	SEQUENCE	333 AA;	37184 MW;	SE0A3D93447680F CRC64;	
Query Match					
41.0%; Score 765; DB 1; Length 333;					
Best Local Similarity 50.8%; Pred.No. 1.4e-53;					
Matches 155; Conservative 49; Mismatches 93; Indels 8; Gaps 9;					
QY	54	FFHFAFFARYKSGS	EEETKKGFGI	PVENLAFIRSTNRKOLSY	FLGINQFADLTWEEFR 113
DB	31	FFHFSNMKQHKTY	-SSVEYNHRLQF	NWNRKIOAHNQBNHT	FKWALNQSDMSFAEIK 89
QY	114	TNRL-GAQNCSATA	GNHRFVDCVL	PVTRDWRQO-IVSP	KDQSCGS-WTSTTTGAL 170
DB	90	HKFLMSEPNCSAT	-KSNYLURGTGP	YPSNMWRKGNVVP	VKNQGCASCWTFSTTGAL 148
QY	171	EAAYTQLTGTSL	--EQLVDCASAF	NFGC-GGLPSQAF	EYVKYNGSIDTEQYPPYLG 227
DB	149	ESAVAIASGKML	SLAEQLVDCAF	AFNNHGCKGLP	SQAFEYLYNKGIMEDSYPIGK 208
QY	228	MGICNFKNQENV	GVGVKVIDSI	INITLCADELKHA	VLGVPRVSVAFVFKGFLNLYKKGVTSSD 287
DB	209	DSSCFRFPQK	AVAFVKVNVNIT	LNDEAAVBEAVAL	NVPVSFAFVTEDFLMYKSGVSSK 268
QY	288	TGCRDPMVHN	AVLAVGYVED	GIPWLKNSNGTNG	DNGYFKMELKNWCGVATCASY 347
DB	269	SCHTPTDKVNH	AVLAVGYG	EGGLLYWTVKNS	WGSQWNGGYFLIRGNMCGLAACASY 328
QY	348	PIVAV	352		
DB	329	PIPQV	333		
RESULT 8					
ID	CATH	HUMAN	STANDARD;	PRT;	335 AA.
AC	P09668;	Q9BUM7;			
DT	01-MAR-1989	(Rel. 10, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Cathepsin H precursor	(EC 3.4.22.16).			
GN	CTSH.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata;	Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini;	Hominidae; Homo.			
OX	NCBI_Taxid=9606;				
FN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver;				
RX	MEDLINE=90067944;	PubMed=2587265;			
RA	Fuchs R.; Gassen H.;				
RT	"Nucleotide sequence of human preprocathepsin H, a lysosomal cysteine				
RL	proteinase.";				
RL	Nucleic Acids Res. 17:9471-9471(1989).				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Eye;				
RX	MEDLINE=22388257;	PubMed=12477932;			
RA	Strausberg R.B.; Feingold E.A.; Grouse L.H.; Derge J.G.,				
RA	Klausner R.D.; Collins F.S.; Wagner L.; Shennan C.W.; Schuler G.D.,				
RA	Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.,				
RA	Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F.,				
RA	Diatchenko L.; Marasina K.; Farmer A.A.; Rubin G.M.; Hong L.,				
RA	Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Schaez T.E.,				
RA	Brownstein M.J.; Usdin T.B.; Tohiyuki S.; Carninci P.; Prange C.,				
RA	Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullaly S.J.,				
RA	Rosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.,				
RA	Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W.,				
RA	Villalón D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.; Sanchez A.,				
RA	Fahney J.; Helton E.; Kettman M.; Madan A.; Rodrigues S.; Sanchez A.,				
RA	Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.,				
RA	Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.,				
RA	Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.,				

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.S., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Mollisacher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vayberg M., Vystetskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:816-820(2000).
 CC -1- INDUCTION: By high salt conditions.
 CC -1- SIMILARITY: Belongs to peptidase family C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D13043; BAA02374.1; -;
 DR EMBL; AC083835; AAG50628.1; -;
 DR PIR; JN0719; JN0719.
 DR HSSP; P00785; 2ACT.
 DR MEROPS; C01.029; -;
 DR InterPro; IPR000118; Granulin.
 DR InterPro; IPR000568; Peptidase_C1.
 DR InterPro; IPR000189; SHPOT_acsite.
 DR Pfam; PF00396; granulin; 1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00277; GRAN; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydrolyase; Thiol protease; Zymogen; Glycoprotein; Signal.
 FT SIGNAL 1 21
 FT PROPEP 22 136
 FT CHAIN 137 462
 FT ACT_SITE 161 161
 FT ACT_SITE 297 297
 FT ACT_SITE 317 317
 FT DISULFID 158 200
 FT DISULFID 192 233
 FT DISULFID 291 342
 FT CARBOHYD 90 90
 FT CARBOHYD 414 414
 SQ SEQUENCE 462 AA; 50966 MW; 4C59955CC895AA58 CRC64;
 Query Match 32.7%; Score 609; DB 1; Length 462;
 Best Local Similarity 38.6%; Pred No. 5.1e-41;
 Matches 140; Conservative 73; Mismatches 112; Indels 38; Gaps 16;
 QY 5 LIFLLFLV-SALSVALAGFEDNPIRSVTORPDSIEPAILGVLSGRHAFHFAPFARY 63
 Db 10 ILFLAMVSSAVDMSIISDEKKGVSSTGGRS---EAEVMSI-----YEAVLVKH 57
 QY 64 GKSYSGEIEIKK--RFGIFVENLAFIRSTNRKDLSTYLGINOFADLTWEFRNRLGAQAQ 121
 Db 58 GKAQSONSLVEKDRFEKFKNLRVDENEKNLSYRLGTFRADLTNDYRSKYLGRAM 117
 QY 122 NCSATAGNHRRF---VDGLVPLVTRDREQGIVSPVKQGGCGS-WTPSTTTGALEAYTOL 177

Db 118 EKKGERSTLRYSRVARVDELDPESIDWRKKGVAEVRKQGGCGSWAFSTIGAVEGINIV 177
 QY 178 TGS--TLSEQLVDCASAFNFC--GGLPSQAFYVYKNGGIDTQTPYPLGWMGICNFK 234
 Db 178 TGLDITLSEQLVDCQTSYNE--CGNGLMDYAFEFIKNGGIDTDKDYPPYKGVDTCDQI 236
 QY 235 QENVGVKVIDSI--NITLGADELKHAUGLVPRVSVAFEV--VKGNLYKKGYYSSDTCCGRD 292
 Db 237 RKNKVVITDSYEDVPTYSSESLKAVA--HQPISTAIETAGRAFLQYDSGIFDG--SCG-- 292
 QY 293 PMDVNHAVLAVGYVEDGIPYMLIKNSWTNGDNGYFMELGKNM-----CGVATCAS 346
 Db 293 -TQLDHGWAAGVGTENGKDYMTVRNSWKGSGESGLRM--ARNIASSSGKCGIAIBPS 349
 QY 347 YPI 349
 Db 350 YPI 352
 RESULT 10
 CATL_SARPE
 ID CATL_SARPE STANDARD; PRT; 339 AA.
 AC Q26636;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15).
 OS Sarcophaga peregrina (flesh fly) (Boettcherisca peregrina).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
 OC Sarcophagidae; Sarcophaga.
 NCBI_TaxID:7386;
 RN NCBI
 RX MEDLINE=94253090; PubMed=8195162;
 RP 224-239 AND 306-319, FUNCTION, SUBCELLULAR LOCATION, AND DEVELOPMENTAL
 RP STAGE;
 RA Homma K.-I., Kurata S., Natori S.;
 RT "Purification, characterization, and cDNA cloning of procathepsin L
 RT from the culture medium of NIH-Sape-4, an embryonic cell line of
 RT sarcophaga peregrina (flesh fly), and its involvement in the
 RT differentiation of imaginal discs.";
 RL J. Biol. Chem. 269:15258-15264(1994).
 CC -1- FUNCTION: Important for the overall degradation of proteins in
 CC lysosomes. Required for differentiation of imaginal
 CC disks.
 CC -1- CATALYTIC ACTIVITY: Specificity close to that of papain. As
 CC compared to cathepsin B, cathepsin L exhibits higher activity
 CC towards protein substrates, but has little activity on L-Arg-Arg-
 CC NMEC, and no peptidyl-dipeptidase activity.
 CC -1- SUBUNIT: Dimer of a heavy and a light chain linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- DEVELOPMENTAL STAGE: Highly expressed during embryonic development
 CC with higher levels in first instar than in third instar.
 CC -1- SIMILARITY: Belongs to peptidase family C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D16533; BAA03970.1; -;
 DR PIR; A53810; A53810.
 DR HSSP; P07711; 1CJL.
 DR MEROPS; C01.092; -;
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHPOT_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PRO0705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
 DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
 DR PROSITE; PS00640; THIOLESTERASE_ASN; 1.
 KW Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal;
 KW Differentiation; Developmental protein.
 FT SIGNAL 1 17
 FT PROPEP 18 121
 FT CHAIN 122 294
 FT CHAIN 122 294
 FT CHAIN 295 339
 FT CHAIN 299 339
 FT ACT_SITE 146 146
 FT ACT_SITE 285 285
 FT ACT_SITE 306 306
 FT DISULFID 143 186
 FT DISULFID 177 219
 FT DISULFID 278 328
 FT CARBOHYD 96 96
 SQ SEQUENCE 339 AA; 37847 MW; 7401F3281A2FDA33 CRC64;
 Query Match 32.4%; Score 603.5; DB 1; Length 339;
 Best Local Similarity 42.3%; Pred. No. 9.7e-41;
 Matches 134; Conservative 48; Mismatches 96; Indels 39; Gaps 11;
 QY 62 RYKGYGSEEEIKRFGIFVENLAFIRSTN---RKLSYTLGINOPADLTWEEF----- 112
 DB 34 QKRYANEVEFRFMKIFENENHKLAKHQLFAGQKYSYKLGKLYADMLHHEFKETMN 93
 QY 113 -----RTNRLGAQNCSTAHGHRFVDGVLPTVRDREQIGVSPVKQSGC 160
 DB 94 GYNHTLRQMRERTGLVATY---IPPAH-----VTVPKSYDREHGAVTGKQDGHG 144
 QY 161 S-WTSTTCALAAATQLTG--STLSEQLVDCASAFNNFGC-GGLPSQAFYVYKNGGI 215
 DB 145 SCWAFSTGALGEGHFRKAGVLVLSQNLVDCSTKYGNNGGNGGLMDNATRYIKDNGGI 204
 QY 217 DTEQYPIYLVGMGICNFKQENVGVKVIDSINITLGADELKHAVGLRVPVSAFVW-KG 275
 DB 205 DTEKSYPEGIDDSCHFNKATIGATDTGFDVPEGDEEKKKAVATMGFPVSAIDASHES 264
 QY 276 FNLYKKGYSSTCGRDMVNVHVLAVGYGV-EDGIPYWLKSNWNTGNDGYPFMEL 334
 DB 265 POLYSEGYNPEEC--DEQNLDHGLVVGVTGDSGMDYLVKSNWNTGWTGEOGYIKVAR 322
 QY 335 GK-NMCGVATCASYPIV 350
 DB 323 NQNNQCGIATASSYPTV 339

RESULT 11
 ID_CYS2_HOMAM STANDARD; PRT; 323 AA.
 AC P25782;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 GN LCP2.
 OS Homarus americanus (American lobster).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
 OC Nephropoidea; Nephropidae; Homarus.
 OX NCBI_TaxID=6706;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Digestive gland;
 RX MEDLINE=92070467; PubMed=1959590;
 RA Laycock M.V., Mackay R.M., di Fruescio M., Gallant J.W.;
 RT "Molecular cloning of three cDNAs that encode cysteine proteinases in
 the digestive gland of the American lobster (Homarus americanus).";
 RL FEBS Lett. 292:115-120(1991).
 CC -!- ENZYME REGULATION: Inhibited by E-64, antipain, leupeptin, heavy

metal ions, iodoacetic acid, dithionitrobenzene, p-hydroxymercuri-
 benzoate; activated by mercaptoethanol and dithiothreitol.
 CC -!- SIMILARITY: Belongs to peptidase family C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X63568; CAA45128.1; --
 DR PIR; S19650; S19650.
 DR HSP; O60911; IPHO.
 DR MEROPS; C01.030; --
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
 DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
 DR PROSITE; PS00640; THIOLESTERASE_ASN; 1.
 KW Hydrolase; Thiol protease; Zymogen; Signal; Multigene family.
 FT SIGNAL 1 15
 FT PROPEP 17 107
 FT CHAIN 108 323
 FT ACT_SITE 131 131
 FT ACT_SITE 270 270
 FT ACT_SITE 290 290
 FT DISULFID 128 171
 FT DISULFID 162 204
 FT DISULFID 263 312
 SQ SEQUENCE 323 AA; 35401 MW; 1AD93FC73C6E9831 CRC64;
 Query Match 31.9%; Score 594.5; DB 1; Length 323;
 Best Local Similarity 43.2%; Pred. No. 4.8e-40;
 Matches 131; Conservative 40; Mismatches 119; Indels 13; Gaps 8;
 QY 59 FARRYKSGSEBEEIKRFGIFVENLAFIRSTNTRK-----DLSYTLGINOPADLTWEEFRT 114
 DB 23 FKGYGRQVDAEDSYRRVIFEQNKQYIEEFNKYENGVEVTFNLANNKFGDMTLEEFNA 82
 QY 115 NRLGAQNCSTAHGHRFVDGVLPTVRDREQIGVSPVKQSGCS-WTSTTCALAA 172
 DB 83 VMKGNTPRSAPVSVFYPKKTGPQATEVDWRTKGAVTPKQGGCGSCWAFSTTGLSG 142
 QY 173 AYTQLTGS--TLSEQLVDCASAFNNFGC-GGLPSQAFYVYKNGIGIDTEQYPIYLVGMG 229
 DB 143 QHFLKTGSLISLAEQQLVDCSRPYGPGCGNGGMMNDADFYIKANNIGIDTEAAYPYEAR 202
 QY 230 ICNFKQENVGVKVIDSINITLGADELKHAVGLRVPVSAFVW-KGFNLYKKGVSST 288
 DB 203 SCRFDSNSVAATCSGHTNIASSETGLQAVRDIGPISVTIDAAKSSFOYSGVYFES 262
 QY 289 CGRDPMDVNVHVLAVGYGVDEGIPYWLKSNWNTGNDGYPFMELGK-NMCGVATCASY 347
 DB 263 C--SPSYLDHVLAVGYGSEGGQDFWLKSNWNTGNDGYPFMELGK-NMCGVATCASY 320
 QY 348 PIV 350
 DB 321 PLV 323
 RESULT 12
 ID_CYS5_DICDI STANDARD; PRT; 344 AA.
 AC P54640;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)

FT ACT SITE 301 BY SIMILARITY.
FT ACT SITE 321 BY SIMILARITY.
FT DISULFID 161 BY SIMILARITY.
FT DISULFID 195 BY SIMILARITY.
FT DISULFID 295 BY SIMILARITY.
FT CARBOHYD 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 471 AA; 50505 MW; 0BF39D33995CEBID CRC64;

Query Match 31.5%; Score 587; DB 1; Length 471;
Best Local Similarity 40.2%; Pred. No. 2.9e-39;
Matches 148; Conservative 55; Mismatches 127; Indels 38; Gaps 15;

Qy 2 AAKLLFLFLV---SALSVALAGFEEDNPIRSYQTPDSEIPAILGVLCGRHAFHAF 58
Db 5 AAAAFLLLVGHRAEDMSIIISNAEHGARGLEEGPTEAE-----ARAAVDLW 54

Qy 59 PARYGKS---YGESEIKKFGIFVENLAFIRSTN-RKUL--SYTLGINQFADLTWEF 112
Db 55 LAENGGS PNALGGEH--RRFLVFDNLKFDVAHNADEGGGFRIGMNRFPADLTNEEF 112

Qy 113 RTNRLGAAQNCATAGHNRHFDVG--LPVTRDREQIVSPVKDQSCGS-WTFSTTGA 169
Db 113 RATFLGAKVAERSAAGERYHDGVEELPESVDREKGAAPVKVQCGSCWAFSAVST 172

Qy 170 LEAYTQLTGS--TLSSQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTQTPYLG 226
Db 173 VESINQLVTGEMITLSQELVECTSCNGSCNGLMADAFDFIKNKGIDTDDYKYA 232

Qy 227 VMGICNFQKQNVGVKVIDSI-NITLGAEDLKHAVGLVRPVSVAFEV-VKGFNLKGVY 284
Db 233 VDGKCDINRENAKVSIDGFEVDQNDKSLQKAVA-HQPVSAIEAGREFQLYHGVF 291

Qy 285 SSDTCGRDPMVNHAVLAGVGVBDGIPYWLKNSWGTNGDNGYFKMELGKNM----CG 340
Db 292 S---GRGCTSLDHGVAVGYGTGDKGDIWYVRNSWGPKNKGSYGVMEKNINVTGKCG 347

Qy 341 VATCASYP 348
Db 348 IAWMASYP 355

RESULT 14
CYSL_HOMAM STANDARD; PRT; 322 AA.
AC P13277;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Digestive cysteine proteinase 1 precursor (EC 3.4.22.-).
GN LCPI
OS Homarus americanus (American lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropoidea; Nephropidae; Homarus.
OX NCBI TaxID=6706;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Digestive gland;
RA MEDLINE=92070467; PubMed=1959590;
RX Laycock M.V., Mackay R.M., di Fruscio M., Gallant J.W.;
RT "Molecular cloning of three cDNAs that encode cysteine proteinases in the digestive gland of the American lobster (Homarus americanus).";
RL FEBS Lett. 292:115-120(1991).
RN [2]
RP SEQUENCE OF 106-133.
RC TISSUE=Digestive juice;
RA MEDLINE=90088376; PubMed=2597115;
RX Laycock M.V., Hirama T., Hasnain S., Watson D., Storer A.C.;
RT "Purification and characterization of a digestive cysteine proteinase from the American lobster (Homarus americanus).";
RL Biochem. J. 263:439-444(1989).
RN [3]
RP ENZYME REGULATION: Inhibited by E-64, antipain, leupeptin, heavy

CC metal ions, iodoacetic acid, dithionitrobenzene, p-hydroxymercuri-
CC benzoate; activated by mercaptoethanol and dithiothreitol.
CC -1- SIMILARITY: Belongs to peptidase family C1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X63567; CA45127.1; -.
DR PIR; S19649; S19649.
DR HSSP; P07711; 1CJL.
DR MEROPS; C01.030; -.
DR InterPro; IPR000668; Peptidase C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; P00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept C1; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CVS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW Hydrolase; Thiol protease; Zymogen; Signal; Multigene family.
FT SIGNAL 1 16
FT PROPEP 17 105
FT CHAIN 106 322
FT ACT_SITE 129 129
FT ACT_SITE 269 269
FT ACT_SITE 289 289
FT DISULFID 126 170
FT DISULFID 160 203
FT DISULFID 262 311
FT CONFLICT 114 114 K -> E (IN REF. 2).
SQ SEQUENCE 322 AA; 35498 MW; EADCD4E3D31291E9 CRC64;

Query Match 31.4%; Score 586; DB 1; Length 322;
Best Local Similarity 42.3%; Pred. No. 2.2e-39;
Matches 131; Conservative 42; Mismatches 109; Indels 28; Gaps 10;

Qy 59 FARRYKSYGSBEEIKKFGIFVENLAFIRSTNRK----DLSTYLGINQFADLTWEFRT 114
Db 23 FKGFGRKVTDLLEERYRLNVLFDNLQYIEEFNKYERGETYTNLAINQFSDMTNEKFA 82

Qy 115 NRLG-----AAQNCATAGHNRHFDVGVLPTVTRDREQIVSPVKDQSCGS-WTFST 166
Db 83 VMKGYKGRPAAVFTSDAPESTE-----DWRITGAVTPVKDQCGSCWAFST 134

Qy 167 TGALEAAVTLTGS--TLSEQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTQTY 222
Db 135 TGIHQHFLKTRLVSLSEQLVDCAGSYNQSCNGWVERALMYVRDNGSVDTESY 194

Qy 223 PYLGVNGICNFQKQNVGVKVIDSI-NITLGAEDLKHAVGLVRPVSVAFEV-VKGFNLK 281
Db 195 PYEARDNTCRFNSNTIGATCTGYVGLAQGSESAKLTATRDIGPISVAIDASHRSFQSYT 254

Qy 282 GYVSSDTCGRDPMVNHAVLAGVGVBDGIPYWLKNSWGTNGDNGYFKMELGK-NMCG 340
Db 255 GYVFPSCSSQLD--HAVLAGVSGEGQDFLWVNSWATSWGSGSYIKMARNNNCG 312

Qy 341 VATCASYPTV 350
Db 313 IATDACYPTV 322

RESULT 15
CATL_DROME STANDARD; PRT; 341 AA.
ID CATL_DROME
AC Q95029; O97431;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15) (Cysteine proteinase 1).
 GN Cpl OR FS(2)50CA OR G6592.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neuroptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TTSUE=Hemocyte;
 RX MEDLINE=97254346; PubMed=909581;
 RA Trysilus Y., Hultmark D.;
 RT "Cysteine proteinase 1 (Cpl), a cathepsin L-like enzyme expressed in
 RT the Drosophila melanogaster haemocyte cell line mbn-2";
 RL Insect Mol. Biol. 6:173-181(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98324856; PubMed=9662479;
 RA Gray Y.H.M., Sved J.A., Preston C.R., Engels W.R.;
 RT "Structure and associated mutational effects of the cysteine
 RT proteinase (Cpl) gene of Drosophila melanogaster";
 RL Insect Mol. Biol. 7:291-293(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Caliniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Wan K.H., Doyle C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Baillet L.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.D., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
 RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zibbe X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2195-2195(2000).
 RN [4]
 RP SEQUENCE OF 34-341 FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=Canton-S;
 RX MEDLINE=95154345; PubMed=7851441;
 RA Matsumoto I., Watanabe H., Abe K., Arai S., Emori Y.;
 RT "A putative digestive cysteine proteinase from Drosophila melanogaster
 RT is predominantly expressed in the embryonic and larval midgut";
 RL Eur. J. Biochem. 227:582-587(1995).

CC -1- FUNCTION: Important for the overall degradation of proteins in
 CC lysosomes. Essential for adult male and female fertility. May play
 CC a role in digestion.
 CC -1- CATALYTIC ACTIVITY: Specificity close to that of papain. As
 CC compared to cathepsin B, cathepsin L exhibits higher activity
 CC towards protein substrates, but has little activity on Z-Arg-Arg-
 CC NHMe, and no peptidyl-dipeptidase activity.
 CC -1- SUBUNIT: Dimer of a heavy and a light chain linked by disulfide
 CC bonds.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- TISSUE SPECIFICITY: In the embryo, predominantly expressed in the
 CC midgut. Also expressed in larval alimentary organs such as
 CC salivary gland and midgut including gastric caeca.
 CC -1- DEVELOPMENTAL STAGE: Expressed in embryo, larva, pupa and adult.
 CC -1- SIMILARITY: Belongs to peptidase family C1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U75652; AAB18345.1; -;
 CC DR EMBL; AF012089; AAB65749.1; -;
 CC DR EMBL; AE003816; AAM68566.1; -;
 CC DR EMBL; D31970; BAA06738.1; ALT_SEQ.
 CC DR HSP; G60911; LFH0.
 CC DR MEROPS; C01.092.1;
 CC FlyBase; FBgn0013770; Cpl.
 CC InterPro; IPR000668; Peptidase C1.
 CC InterPro; IPR000169; SHprot_accsite.
 CC Pfam; PF00112; Peptidase C1; 1.
 CC PRINTS; PR00705; PAPAIN.
 CC ProDom; PD000158; Peptidase C1; 1.
 CC SMART; SM00645; Pept C1; 1.
 CC PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 CC PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
 CC PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 CC Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal;
 CC Digestion; Developmental protein.
 CC SIGNAL
 CC FT PROPEP 1 18 POTENTIAL.
 CC FT PROPEP 19 123 ACTIVATION PEPTIDE.
 CC FT CHAIN 124 296 CATHEPSIN L HEAVY CHAIN.
 CC FT PROPEP 297 299
 CC FT CHAIN 300 341
 CC FT ACT_SITE 148 148 CATHEPSIN L LIGHT CHAIN.
 CC FT ACT_SITE 287 287 BY SIMILARITY.
 CC FT ACT_SITE 308 308 BY SIMILARITY.
 CC FT DISULFID 145 188 BY SIMILARITY.
 CC FT DISULFID 179 221 BY SIMILARITY.
 CC FT DISULFID 280 330 INTERCHAIN (BY SIMILARITY).
 CC FT CARBOHYD 97 97 N-LINKED (GLCNAC...) (POTENTIAL).
 CC FT CONFLICT 198 198 R -> P (IN REF. 4).
 CC FT CONFLICT 225 227 KGT -> RAO (IN REF. 4).
 CC FT CONFLICT 247 251 AEAFA -> PEPVP (IN REF. 4).
 CC FT CONFLICT 335 335 A -> P (IN REF. 4).
 CC SQ SEQUENCE 341 AA; D18382396ACEID59 CRC64;
 CC
 CC Query Match 31.28; Score 581; DB 1; Length 341;
 CC Best Local Similarity 42.48; Pred No. 6e-39;
 CC Matches 134; Conservative 45; Mismatches 107; Indels 30; Gaps 11;
 CC
 CC QY 59 FARRYGKSYGSEBEIKRFGFVFNIAFTSTNRK-----DLSTYTGINQFADLTWERPRT 114
 CC 32 FKLEHRKNYQDETEERFLKIFNENKHKAKNORFAEGKVSFKLVANKYADLLHHEFRQ 91
 CC 115 NRLGAQNCATAGNHRFVD-----GV-----LPVTRDRBQGVVSPVQGGSCGS 161
 CC 92 LMWG-----FNVLHKLQRAADESFKGVFTISPAHVTLPKSDVDRTKGAVTAKQGHGS 147
 CC 162 -WTSTTTCALAAAYTQLTG--STLSQQQLVDCASAFNFGC--GGLPSQAFYKVGNGID 217

Db	148	CWAFSSTGALSGQHRKSGVLVLSLSEQLVDCSTKYGNNGCNGGLMDNAPFYIKONGGID	207
Qy	218	TEQTPYPLGVMGICNFKQENTGVKVIDSINITLGADELKHAUCLVRPVSVAFEV- KGP	276
Db	208	TEKSYFYEAI DDSCHFNKGTGATDRGFTDIPOGDEKQKMAEAVATVGPVSVAIDASHESF	267
Qy	277	NLYKKGVYSSDTCCGRDPMDVNHAVLAVGYGV- EDGIPYWLKNSWGTNMGDNGYFKWELG	335
Db	268	QFYSEGVYNEPQC--DAQNLDHGLVVGFGTDESGEDYWLVKNSWGTWGDGKGFIKMLRN	325
Qy	336	K-NMCGVATCASYPIV	350
Db	326	KENQGTIASASSYPLV	341

Search completed: March 29, 2004, 22:12:14
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:02:04 ; Search time 79 Seconds

(without alignments)
1405.853 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865
Sequence: 1 MAAKLFFLLFLVLSVAL.....ELGKMGVATCASYPIVAV 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

1: sp_arbcea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1310.5	70.3	360	10 Q43705	Q43705 zea mays (m
2	1297	69.5	359	10 Q8X178	Q8X178 brassica ol
3	1286	69.0	360	10 Q9XR12	Q9XR12 nicotiana t
4	1282	68.7	359	10 Q9M4E5	Q9M4E5 lolium mult
5	1279.5	68.6	358	10 Q9L183	Q9L183 arabidopsis
6	1279.5	68.3	358	10 Q8H166	Q8H166 arabidopsis
7	1271	68.2	360	10 Q84YH8	Q84YH8 nicotiana t
8	1265.5	67.9	358	10 Q8WQ9	Q8WQ9 arabidopsis
9	1262.5	67.7	358	10 Q40877	Q40877 petunia rpb
10	1261	67.6	358	10 Q50002	Q50002 prunus arne
11	1254.5	67.3	350	10 Q41057	Q41057 pisum sativ
12	1252.5	67.2	350	10 Q949A2	Q949A2 pisum sativ
13	1207.5	64.7	377	10 Q9W3E7	Q9W3E7 arabidopsis
14	763	40.9	333	11 Q922Q7	Q922Q7 mus musculu
15	763	40.9	335	4 Q96NV6	Q96NV6 homo sapien
16	722	38.7	178	10 Q9SDN1	Q9SDN1 prunus dulc

17	679	36.4	366	5 Q86G98	Q86G98 oxytricha t
18	661.5	35.5	332	5 Q7YW75	Q7YW75 rhipicephal
19	656.5	35.2	326	5 Q24940	Q24940 fasciola gi
20	655.5	35.1	326	5 Q3NGW4	Q3NGW4 fasciola gi
21	651.5	34.9	326	5 Q8MUT6	Q8MUT6 fasciola gi
22	648	34.7	335	5 Q7YW74	Q7YW74 rhipicephal
23	647.5	34.7	326	5 Q9XYL8	Q9XYL8 fasciola gi
24	644.5	34.6	326	5 Q24941	Q24941 fasciola he
25	642.5	34.5	326	5 Q721E9	Q721E9 fasciola he
26	641.5	34.4	326	5 Q9NB30	Q9NB30 fasciola he
27	637	34.2	462	10 Q93XQ9	Q93XQ9 ipomoea bat
28	637	34.2	485	10 Q8W180	Q8W180 brassica ol
29	634.5	34.0	326	5 Q9NGW3	Q9NGW3 fasciola gi
30	632.5	33.9	332	5 Q9NHBS	Q9NHBS boophilus m
31	631.5	33.9	311	5 Q9GRW5	Q9GRW5 fasciola he
32	630.5	33.8	310	5 Q96710	Q96710 fasciola he
33	629.5	33.8	311	5 Q8T529	Q8T529 fasciola he
34	627	33.6	325	5 Q26360	Q26360 fasciola sp
35	625.5	33.5	324	5 Q8T0X0	Q8T0X0 fasciola he
36	624.5	33.5	324	5 Q86GJ2	Q86GJ2 hydra atten
37	621.5	33.3	338	5 Q9Y0X2	Q9Y0X2 artemia san
38	621.5	33.3	461	10 Q84M29	Q84M29 helianthus
39	616.5	33.1	351	5 Q8IT42	Q8IT42 theromyzon
40	616	33.0	374	10 Q24137	Q24137 nicotiana t
41	615	33.0	463	10 Q9FWH8	Q9FWH8 arabidopsis
42	613.5	32.9	338	5 Q46030	Q46030 sitophilus
43	612	32.8	374	10 Q84TH7	Q84TH7 nicotiana t
44	611.5	32.8	306	5 Q7Z0G8	Q7Z0G8 metapenaeus
45	611.5	32.8	322	5 Q7Z0G9	Q7Z0G9 metapenaeus

ALIGNMENTS

RESULT 1

Q43705 PRELIMINARY; PRT; 360 AA.

AC Q43705; DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Cysteine protease precursor.
GN SEEL.
OS zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. ES; TISSUE=Leaf;
RX MEDLINE=97422404; PubMed=9278172;
RA Griffiths C.M., Hosken S.E., Oliver D., Chojek A.J.S., Thomas H.;
RT "Sequencing, expression pattern and RFLP mapping of a senescence-
enhanced cDNA from Zea mays with high homology to oryzain gamma and
aleurain";
RL Plant Mol. Biol. 34:815-821(1997).
DR EMBL; X99936; CAA68192.1; -;
DR HSP; O46427; 8PCH.
DR GO; GO:0004187; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE CIS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE HIS; 1.
DR HydroLase; Protease; Signal; Thiol protease.
FT SIGNAL 29 34 POTENTIAL.
FT CHAIN 143 360 POTENTIAL.

Db 13 LALILAASTAESIGFDESNPIRMVSDRLREVEESVQVLQGSRRHVSFARFAHRYGKRYE 72

Qy 69 SBEIEIKRGEIEMELAFIRSTNRKDLSTYTGINQFADLTWEEPRTRLNRLGAQNCSATAH 128

Db 73 NAEEMKLRFSIFKENLDLIRSTNKKGLSYKLVGNQFADMTWQEFQRTKLGAQNCSATLK 132

Qy 129 GNRHFRVDGVLPTVRDRQGGIVSPVKDOGSCGS--WTSTTTGALEAAAYTQTGS--TISEQ 195

Db 133 GTHKUTGEALPTKDRSDGIVSPVKDOGCGSCSWTFTTGALEAAAYHQAFGKGISLSEQ 192

Qy 186 QLVDCASAFNNFGC--GGLPSQAFEVYKNGGIDTEQTPYLVGNVGICNFKQENVGVKVID 244

Db 193 QLVDCAGAFNNYCGNGLPSQAFEVYKNSGGLDTEREAYPTGEGDTCKYSAENVGVVELD 312

Qy 245 STNITLGADELKHAVLGVRPVSVAFEVVKGNLYKGVYSSDTCGDRDPMDVHVLAVG 304

Db 253 SVNITLGADELKHAVLGVRPVSVAFEVVIHFSFRIYKGVYSDSHCGGTIPMDVHVLAVG 312

Qy 305 YGVEDGIPWLIKNSGNGWNGDGYFKMELKKNKCGVATCASYPIVA 351

Db 313 YGTEDGVPWLIKNSGADWDGKGYFKMGMKNGMCGIATCASYPVVA 359

RESULT 3

Q9LR12 PRELIMINARY; PRT; 360 AA.

ID Q9LR12 AC Q9LR12

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Cysteine protease precursor.

GN NTPC-23.

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC Lamiales; Solanales; Solanaceae; Nicotiana.

OX NCBI_TaxId=4097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Samsun NN; TISSUE=Leaf;

EX MEDLINE=21039373; PubMed=11198425;

RA Ueda T, Seo S, Ohashi Y., Hashimoto J.;

RT "circadian and senescence-enhanced expression of a tobacco cysteine

RT protease gene.;"

RL Plant Mol. Biol. 44:649-657(2000).

EN ENBL; AB032168; BAA96501.1; -.

DR HSSP; P07711; 1CJL.

DR MEROPS; C01.041; -.

DR GO; GO:0004507; P:cysteine-type endopeptidase activity; IEA.

DR GO; GO:006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR000668; Peptidase_C1.

DR InterPro; IPR000169; Shprot_acsite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept C1; 1.

DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.

DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.

DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.

KM Hydrolase; Protease; Signal; Thiol protease.

FT SIGNAL 31 36

FT CHAIN 143 360

SEQUENCE 360 AA; 39199 MW; BFAEC40F2D3CD957 CRC64;

Query Match 69.0%; Score 1286; DB 10; Length 360;

Best Local Similarity 69.3%; Pred. No. 7.7e-101;

[illegible]

RN SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kaniya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT Arabidopsis Full Length cDNA Clones.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A.
 RP Stracke R., Palme K.;
 RA "Signal Peptide Selection derived cDNAs from Arabidopsis thaliana
 RT leaves and guard cells";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF231883; AAF43041.1; -
 DR EMBL; AF011483; BAB08221.1; -
 DR EMBL; AF360273; AAK25983.1; -
 DR EMBL; AY089662; AAM66984.1; -
 DR EMBL; BT000673; AAN31819.1; -
 DR EMBL; BT000676; AAN31822.1; -
 DR EMBL; AF083703; AAN60262.1; -
 DR PIR; P00650; P00650.
 DR HSP; P07711; 1CJL.
 DR MEROPS; C01.041; -
 DR CO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR CO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
 KW Hypothetical protein; Hydrolase; Protease; Thiol protease.
 SQ SEQUENCE 358 AA; 38959 MW; 3B610AB85F81C31D CRC64;

 Query Match 68.6%; Score 1279.5; DB 10; Length 358;
 Best Local Similarity 67.0%; Pred. No. 2.7e-100;
 Matches 240; Conservative 39; Mismatches 72; Indels 7; Gaps 4;

 QY 1 MAALKLF---FLFLVLSALVALAGFEEDNPISVTRQDPSIEPAILGVLSGCRHAFHFA 57
 DB 1 MSATILSSVVLVVLVAASAAANIGFDESPIRMSVDGLREVESVSLQGSRHVLSFA 60

 QY 58 RFARYGKYSGBEIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFRTNRL 117
 DB 61 RFTHRYGKYYQNVBEEMKLRFSIFKENDLIRSNKKGLSYKLVGNQFADLTWQEFQTKL 120

 QY 118 GAAQNCSTAHGNHRFVDGVLPTVRDWRQGIIVSPVKDQSGCS-WTFSTTGALAAAYTQ 176
 DB 121 GAAQNCSTALTKGSHKVTAAALPETKQWREDGIIVSPVKDQSGCSWTFSTTGALAAAYHQ 180

 QY 177 LTGS--TLSEQQLVDCASAFNNFC--GGLPSQAFVYKNGGIDTQTPYPLGVNGICNF 233
 DB 181 AFGKGISLSEQQLVDCAGAFNNYGCNGSLPSQAFVYKNGGLDTEKAYPTTGKDETCKF 240

 QY 234 KOENVGVKVIDISINITLGADELKHAHLVLRPVSVAFEVVKFNLYKKGVSSTDCGRDP 293
 DB 181 AFGKGISLSEQQLVDCAGAFNNYGCNGSLPSQAFVYKNGGLDTEKAYPTTGKDETCKF 240

 QY 294 KOENVGVKVIDISINITLGADELKHAHLVLRPVSVAFEVVKFNLYKKGVSSTDCGRDP 293
 DB 241 SAENVGVQLVNSVNIITLGADELKHAHLVLRPVSVAFEVVKFNLYKKGVSSTDCGRDP 300

 QY 294 MDVNHAVLAVGVYVEDGIPYWLKNSGWTWNGDGFYKMLGKNGMCGVATCASYPIVA 351
 DB 301 MDVNHAVLAVGVYVEDGIPYWLKNSGWTWNGDGFYKMLGKNGMCGVATCASYPIVA 358

 RESULT 6
 Q8H166 PRELIMINARY; PRT; 358 AA.
 AC Q8H166;

DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative cysteine proteinase AAP.
 GN AT5G60360.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kaniya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT Arabidopsis Full Length cDNA Clones.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BT000674; AAN31820.1; -
 DR CO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR CO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acSite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
 SQ SEQUENCE 358 AA; 39007 MW; 2524405745617F17 CRC64;

 Query Match 68.3%; Score 1274.5; DB 10; Length 358;
 Best Local Similarity 66.8%; Pred. No. 7.2e-100;
 Matches 239; Conservative 39; Mismatches 73; Indels 7; Gaps 4;

 QY 1 MAALKLF---FLFLVLSALVALAGFEEDNPISVTRQDPSIEPAILGVLSGCRHAFHFA 57
 DB 1 MSATILSSVVLVVLVAASAAANIGFDESPIRMSVDGLREVESVSLQGSRHVLSFA 60

 QY 58 RFARYGKYSGBEIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFRTNRL 117
 DB 61 RFTHRYGKYYQNVBEEMKLRFSIFKENDLIRSNKKGLSYKLVGNQFADLTWQEFQTKL 120

 QY 118 GAAQNCSTAHGNHRFVDGVLPTVRDWRQGIIVSPVKDQSGCS-WTFSTTGALAAAYTQ 176
 DB 121 GAAQNCSTALTKGSHKVTAAALPETKQWREDGIIVSPVKDQSGCSWTFSTTGALAAAYHQ 180

 QY 177 LTGS--TLSEQQLVDCASAFNNFC--GGLPSQAFVYKNGGIDTQTPYPLGVNGICNF 233
 DB 181 AFGKGISLSEQQLVDCAGAFNNYGCNGSLPSQAFVYKNGGLDTEKAYPTTGKDETCKF 240

 QY 234 KOENVGVKVIDISINITLGADELKHAHLVLRPVSVAFEVVKFNLYKKGVSSTDCGRDP 293
 DB 241 SAENVGVQLVNSVNIITLGADELKHAHLVLRPVSVAFEVVKFNLYKKGVSSTDCGRDP 300

 QY 294 MDVNHAVLAVGVYVEDGIPYWLKNSGWTWNGDGFYKMLGKNGMCGVATCASYPIVA 351
 DB 301 MDVNHAVLAVGVYVEDGIPYWLKNSGWTWNGDGFYKMLGKNGMCGVATCASYPIVA 358

 RESULT 7
 Q84YH8 PRELIMINARY; PRT; 360 AA.
 AC Q84YH8;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE NTP23-like cysteine proteinase.
 OS Nicotiana tabacum (Common tobacco).


```
DR InterPro; IPR000169; Shprot_acsite.  
DR Pfam; PF00112; Peptidase C1; 1.  
DR PRINTS; PR00705; PAPAIN.  
DR ProDom; PD000158; Peptidase_C1; 1.  
DR SMART; SM00645; Pept C1; 1.  
DR PROSITE; PS00038; HLH_1; 1.  
DR PROSITE; PS00640; THIOI_PROTEASE ASN; 1.  
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.  
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.  
KW Hydrolase; Proteaser; Thiol protease..  
SQ SEQUENCE 358 AA; 39309 MW; C98F78793B002554 CRC64;  
  
Query Match          67.6%; Score 1261; DB 10; Length 358;  
Best Local Similarity   68.0%; Pred.No.1e-98;  
Matches 236; Conservative    35; Mismatches 72; Indels 4; Gaps 3  
  
QY      9 LLFLVSLALSVALAAGFEEDNPIRSTVTQRDPDIEPAILLGVLSGRHFAPFAFRFYKSGSYG 68  
         :|:|:|:|:|:|:~::~||::||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:  
Db     12 LVLVAISCGAAASPFDSNPIRLVSDGLRELEQQVVQLNSRRALFFARFAHYRGKKYE 71  
        |::|:|:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:  
QY      69 SEEETKEFGFI PVENLAFISTWNRKDLSYLTLGNQADLTWEFRPNRNLGAONCSATAH 128  
         ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
Db     72 SVEEMKLRIEIFSNNKLLIRTNKNKGIDLYTLANVRPADWSWEEFKRKRLGAQAQC SATTK 131  
         |:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
QY     129 GNHRVDVGVLPTVRDMZREQIGVPVKDQGSCGS--WTFTTGALRAAYTQTGCS--TLSEQ 185  
         ::||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
Db     132 GSHELTDAVLPESKXMRREGIVTPVKDDQHGGCSCWTFSTTTGAEAAVVOAFRKQISLSEQ 191  
         ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
QY     186 QLVDCASAFNNFGC-GGLPSOAIFYVNKNGIDITECTYPVLGMVGTCNFPKQNMGVKKVID 244  
         ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
Db     192 QLVDCAGAFNNFGCHGGLGPSOAFEIKYNGLDTEAARYPVGTGDACKESAVENVGVQLD 251  
         ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
QY     245 SINITLGAEDELKHANGLVFPVSAREVVKFNLYKGYVSSDTCCRDDPMDVNHAVLAG 304  
         |:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
Db     252 SVNITLGDQEQLRHAVA FVPVSVAVQVVKFSFRYYKSGETYTSDTCSSPMDNVNHAVLAG 311  
         ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
QY     305 YGEVDGI PYWLINKSNWGNTNGDNRYFMELGKMCMCVATCASYPIVA 351  
         ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
Db     312 YGEEGVGFPLINKSWGMSGNDNGYFKMFPGKMKMCGVATCASYPIVA 358  
         ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||  
  
RESULT 11  
Q41057 PRELIMINARY; PRT; 350 AA.  
AC AC Q41057  
DT DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE DE Cysteine protease.  
OS Os Pisum sativum (garden pea).  
OC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC OC eurosoids I; Fabales; Papaceae; Papilionoideae; Viciales; pisum.  
RN RN NCBI_TaxId=3988;  
RX RX [1]_TaxId=3988;  
RP RP SEQUENCE FROM N.A.  
RC RC STRAIN=Peacham First;  
RX MEDLINE=96350414; PubMed=9765223;  
RA Jones C.G.; Tucker G.A., Lycett G.W.;  
RT "Pattern of expression and characteristics of a cysteine protease cDNA  
from pea (Pisum sativum L.).";  
RL Biochim. Biophys. Acta 1296:13-15(1996).  
RD EMBL; Z68291; CAAG2583.1; -;  
DR PIR; S71923; S71923.  
DR HSP; O46427; 8PCH.  
DR MEROPS; COI 041;  
DR GO; GO:0004137; F:cysteine-type endopeptidase activity; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000668; Peptidase_C1.  
DR InerPro; IPR000169; Shprot_acsite.  
DR Pfam; PF00112; Peptidase C1; 1.  
DR PRINTS; PR00705; PAPAIN.  
DR ProDom; PD000158; Peptidase_C1; 1.
```

DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOLE PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOLE PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOLE PROTEASE_HIS; 1.
 KW Hydrolyase; Protease; Thiol protease.
 SQ SEQUENCE 350 AA; 38575 MW; 0D7F8F2C51531893 CRC64;

Query Match 67.3%; Score 1254.5; DB 10; Length 350;
 Best Local Similarity 67.2%; Pred. No. 3.5e-98;
 Matches 240; Conservative 39; Mismatches 65; Indels 13; Gaps 6;

QY 1 MAAKLLFFLLFLYSALVALAG--FEEDNPIRSVTORPDSIEPAILGLVLSCHRAHFA 58
 DB 1 MAQWSLLIVLCFVAS---AAAGSFHDSNPIRMVS---DVEBQLLQVIGESHAVSFAR 53
 QY 59 FARRYKSGYSEEEIKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFRNRLG 118
 DB 54 FANRYGKRYDSVDENKLRKIFSENLELRISNKRRLSYKLVGNHFDWTFEERSHRLG 113
 QY 119 AAQNCSTAHGNERFVDGVLPTVRDREOGIIVSVKDGSCGS-WTFSTTGALSAAYTOL 177
 DB 114 AAQNCSTATLGNHKTIDANLPDEKWRKEGIVSGVKDQSGSCWTFSTTGALSAAYTOL 173
 QY 178 TGS--TLSEQLQVDCASAFNFGC--GGLPSQAFYVYKNGGIDTEQTPYLVGMGICNFK 234
 DB 174 FGNISLSEQLQVDCAGAFNFGC--GGLPSQAFYVYKNGGIDTEQTPYLVGMGICNFK 233
 QY 235 QENVGVKVIDSINITLGADELKHAIVLRPVSAFVGVKGNLYKKGIVSSDTGCRDPM 294
 DB 234 SEHAVKVLGVSNNITLGADELKHAIFARPVSAFVGVKGNLYKKGIVSSDTGCRDPM 293
 QY 295 DVNHAVLAGVGYVEDGIPYWLKNSGNTWNGDGYFKMELGKMGVATCASYPIVA 351
 DB 294 DVNHAVLAGVGYGIEDGIPYWLKNSGNTWNGDGYFKMELGKMGVATCASYPIVA 350

RESULT 12
 Q949A2 ID Q949A2 PRELIMINARY; PRT; 350 AA.
 AC Q949A2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Early leaf senescence abundant cysteine protease.
 GN ELSA.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Messire;
 RA Pic E.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ278699; CAC41636.1; -.
 DR MEROPS; C01.041; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR Pfam; PF00112; SHprot_acsite.
 DR PRINTS; PR00705; PAPAIN.
 DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOLE PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOLE PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOLE PROTEASE_HIS; 1.
 KW Hydrolyase; Protease; Thiol protease.
 SQ SEQUENCE 350 AA; 38575 MW; 0CBFF2F87B61892 CRC64;

Query Match 67.2%; Score 1252.5; DB 10; Length 350;
 Best Local Similarity 66.9%; Pred. No. 5.2e-98;

QY 1 MAAKLLFFLLFLYSALVALAG--FEEDNPIRSVTORPDSIEPAILGLVLSCHRAHFA 57
 DB 1 MAQWSLLIVLCFVAS---AAAGSFHDSNPIRMVS---DVEBQLLQVIGESHAVSFAR 53
 QY 59 FARRYKSGYSEEEIKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFRNRLG 118
 DB 54 FANRYGKRYDSVDENKLRKIFSENLELRISNKRRLSYKLVGNHFDWTFEERSHRLG 113
 QY 119 AAQNCSTAHGNERFVDGVLPTVRDREOGIIVSVKDGSCGS-WTFSTTGALSAAYTOL 177
 DB 114 AAQNCSTATLGNHKTIDANLPDEKWRKEGIVSGVKDQSGSCWTFSTTGALSAAYTOL 173
 QY 178 TGS--TLSEQLQVDCASAFNFGC--GGLPSQAFYVYKNGGIDTEQTPYLVGMGICNFK 234
 DB 174 FGNISLSEQLQVDCAGAFNFGC--GGLPSQAFYVYKNGGIDTEQTPYLVGMGICNFK 233
 QY 235 QENVGVKVIDSINITLGADELKHAIVLRPVSAFVGVKGNLYKKGIVSSDTGCRDPM 294
 DB 234 SEHAVKVLGVSNNITLGADELKHAIFARPVSAFVGVKGNLYKKGIVSSDTGCRDPM 293
 QY 295 DVNHAVLAGVGYVEDGIPYWLKNSGNTWNGDGYFKMELGKMGVATCASYPIVA 351
 DB 294 DVNHAVLAGVGYGIEDGIPYWLKNSGNTWNGDGYFKMELGKMGVATCASYPIVA 350

RESULT 13
 Q9M3E7 ID Q9M3E7 PRELIMINARY; PRT; 377 AA.
 AC Q9M3E7;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cysteine protease-like protein.
 GN F18N11.70.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jordan N., Bangert S., Wiedemann R., Voss H., Unsel M., Mewes H.W.,
 Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.,
 Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL132953; CAB72480.1; -.
 DR PIR; T47471; T47471.
 DR HSP; O46427; SPCH.
 DR MEROPS; C01.041; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR Pfam; PF00112; SHprot_acsite.
 DR PRINTS; PR00705; PAPAIN.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOLE PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOLE PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOLE PROTEASE_HIS; 1.
 KW Hydrolyase; Protease; Thiol protease.
 SQ SEQUENCE 377 AA; 41835 MW; DDC2082A6739A180 CRC64;

Query Match 64.7%; Score 1207.5; DB 10; Length 377;
 Best Local Similarity 65.9%; Pred. No. 3.8e-94;
 Matches 228; Conservative 37; Mismatches 74; Indels 7; Gaps 4;

QY 1 MAAKLLFFLLFLYSALVALAG--FEEDNPIRSVTORPDSIEPAILGLVLSCHRAHFA 57

	Query Match	40.9%	Score 763	DB 11	Length 333
	Best Local Similarity	50.8%	Pred. No. 1.9e-56		
	Matches 155	Conservative 49	Mismatches 93	Indels 8	Gaps 7
Qy	54	PHARPARRKXGKSYGSEBEIKKRGIFVENLAFTIRNKKDLSYTLGINOFADLTWEER	113		
Db	31	PHFKSMWKQHKYTSSEVYENRLOMPANNRKIOAHNRNHTFKOALNQSDMSFAEK	89		
Qy	114	TNRL-GAAQNCATAGHNHRFVDGVLPTDRDREQ-IVSPVKDQSGCS-WTFSTTGAL	170		
Db	90	HKFLWSPQNCAT-KSNVLRGTGTPSSMDRKKGVSPVINGQACGSCWTFSTTGAL	148		
Qy	171	EAAVTQLTGSTLS-EQQLVDCASAFNNFC-GGLPSQAFYVTKNGIDITQTYPLVG	227		
Db	149	ESAVATSGKMLSLAEQQLVPCAOAFNNHCKGGLPSQAFYILYKNGIMEESUSPYIGK	208		

Q96NY6

Q: E4 EUEEPEAEPYCKVCYGCEPPTKVDECTEVLENT.AETPSTAPKRI.SVTI.CINOFADT.TWPEPP 113

[illegible]

Search completed: March 29, 2004, 22:15:12
Job time : 81 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 22:22:05 ; Search time 2371 Seconds
(without alignments)
4433.356 Million cell updates/sec

Title: US-10-087-714-2
Perfect score: 1865
Sequence: 1 MAAKLFFLLFLVSALSVAL.....ELGKMGVATCATSYPIVAV 352

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Ddelop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931050276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10087714/runat_26032004_113453_946/app.query.fasta_1.519
-DB=EST_QPWT-fasCap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087714.qcgn_1_1_2135@runat_26032004_113453_946 -ICPU=3
-NO.MMAP -LARGQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1321.5	70.9	1532	11	AY106223	Zea mays
2	1097.5	58.8	989	14	CK263045	CK263045
3	1038	55.7	985	14	CK287181	CK287181
4	1035	55.5	829	14	CB341855	CB341855
5	1034	55.4	977	14	CK295860	CK295860
6	1022.5	54.8	1005	14	CF212812	CF212812
7	1014.5	54.4	936	14	CK268548	CK268548
8	1004.5	53.9	920	14	CK257583	CK257583
9	988	53.0	748	14	CB982399	CB982399
10	987	52.9	814	12	BG594735	BG594735
11	981	52.6	965	14	CK294622	CK294622
12	975.5	52.3	731	13	BQ281381	BQ281381
13	970	52.0	768	14	CB892658	CB892658
14	967.5	51.9	794	10	BE19255	BE19255
15	957	51.3	720	10	BE131652	BE131652
16	944	50.6	924	14	CK294694	CK294694
17	943	50.6	930	14	CK295057	CK295057
18	941	50.5	946	14	CK285936	CK285936
19	938	50.3	848	14	CF510505	CF510505
20	934	50.1	886	14	CF510584	CF510584
21	925	49.6	901	14	CK298522	CK298522
22	919.5	49.3	910	13	BQ797116	BQ797116
23	919	49.3	701	14	CB894684	CB894684
24	918.5	49.2	1230	10	BE422216	BE422216
25	915.5	49.1	852	14	CK263154	CK263154
26	915	49.1	760	10	BF267048	BF267048
27	915	49.1	888	14	CK292918	CK292918
28	915	49.1	911	14	CK285268	CK285268
29	912	48.9	851	14	CB891143	CB891143
30	910	48.8	653	14	CB917819	CB917819
31	910	48.8	899	14	CK293298	CK293298
32	909.5	48.8	709	12	BG155091	BG155091
33	901	48.3	755	12	BG645271	BG645271
34	891	47.8	894	13	BU039998	BU039998
35	885	47.5	703	12	B1263594	B1263594
36	884.5	47.4	827	12	BG645454	BG645454
37	883.5	47.4	686	14	CF477066	CF477066
38	880.5	47.2	842	14	CF517076	CF517076
39	880	47.2	648	13	BQ625193	BQ625193
40	879.5	47.2	844	14	CF518186	CF518186
41	876	47.0	755	14	CB981790	CB981790
42	875	46.9	635	13	BU039803	BU039803
43	874.5	46.9	639	10	AM158820	AM158820
44	874.5	46.9	665	10	AW399964	AW399964
45	874	46.9	710	12	B1267420	B1267420

ALIGNMENTS

RESULT 1
AY106223
LOCUS Zea mays PC0114759 mRNA sequence.
DEFINITION Zea mays
ACCESSION AY106223
VERSION AY106223.1 GI:21209301
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1532)

/tissue type="abiotic stress treated leaf and root tissue"
 /lab host="DH10B-Tona"
 /clone_lib="potato abiotic stress cDNA library"
 /note="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: Solum tuberosum var. Kennebec plants were
 grown from cuttings on a 16hr light/8 hr dark cycle at 25
 C for 3-4 weeks. Abiotic stress conditions were applied to
 four separate sets of plants. Set 1 involved saturation of
 the soil with 150 mM NaCl and tissues were harvested at
 following application of the salt stress (leaves: 2hr,
 6hr, 12hr, 1d, 2d, and 4d; roots: 2hr, 6hr, 12hr, 1d,
 2d, and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d,
 2d. Set 4 were grown under the standard conditions and
 then were heat stressed by placement at 35 C. Heat
 stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
 2d and 4d and heat-stressed roots were harvested at 6 hr,
 12 hr, 1 d, and 4d. RNA was isolated from all tissues and
 equal RNA from each tissue and stress was pooled to
 construct the cDNA library. RNA sample."

ORIGIN

Alignment Scores:
 Pred. No.: 5,76e-120 Length: 989
 Score: 1097.50 Matches: 220
 Percent Similarity: 77.30% Conservative: 32
 Best Local Similarity: 77.48% Mismatches: 65
 Query Match: 58.85% Indels: 9
 DB: 14 Gaps: 6

US-10-087-714-2 (1-352) x CK263045 (1-989)

Qy 6 LeuPheLeuLeuPheLeuValSerAlaLeu---SerValAlaLeuAlaGly----- 22
 Db 15 CTCTCACTCTATTGGTTCGTGCGCGGCTTTTGTCTGCTGCACTTGCCTGCGGACCGGCG 74
 Qy 23 ---PheGluGluAspAsnProIleArgSerValThrGlnArgProAspSerIleGluPro 41
 Db 75 ACTTTCGCCATGAGATCCGATCAGGCAAGTGTAGTT---TCCGAGAGCTGGAGAAC 131
 Qy 42 AlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArg 61
 Db 132 GGAATTCCTCAAGTCGTGCGCCAGACTCGCAATGCTCTCTCTCTGCTTCTGCTATC 191
 Qy 62 ArgTyrGlyLysSerTyrGlySerGluGluIleLysLysArgPheGlyIlePheVal 81
 Db 192 AGGCATCGGAAGAGTACGAGTCGTTGAGGAGATCAAGCAAGGTTCCAGATATTTTG 251
 Qy 82 GluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIle 101
 Db 252 GACAATCTGAAGATGATCCGATCGCATAACAGCAAGGACTATACAACTCGGTGC 311
 Qy 102 AsnGlnPheAlaAspLeuThrTrpGluPheArgThrAsnArgLeuGlyAlaAlaGln 121
 Db 312 AATGAGTTTACCCACTTAACATGGATGAGTTCCTGAGGATCAAGCAAGTTCGGGCACTCAA 371
 Qy 122 AsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThr 141
 Db 372 AACTGTTCTGCCACTACAAGGCGCATCTCAAGCTAACTAAGCTTTCTGCCAGGAGC 431
 Qy 142 ArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer 161
 Db 432 AAGGACTCGAGGAAGATGATTGTTAGCCAGTGAAGGCAAGGCAAGTTCGGGATC 491
 Qy 162 ---TrpThrPheSerThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer 180
 Db 492 TGTGACATTCAGCACTCTGTGTCATAGAGGCGAGCATATGCCCAAGCATTTGGGAAG 551

Qy 181 -----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnGlnPheGly 198
 Db 552 GGAATCTCTCTCTGACAGCAGCAGCTTGTGGACTGTGCTGGAGCTTTTAAATAACTTTGGC 611
 Qy 199 Cys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAsp 217
 Db 612 TGCATGGGGGGTGGCCATCACAGCCCTTGGTACATTAATTCATGGTGGTCTTGAC 671
 Qy 218 ThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsn 237
 Db 672 ACTGAAGAGCATATCCATACACCGCAAGATGCATATGTAATTTCTCAACAGCAAT 731
 Qy 238 ValGlyValLysValLysSerIleAsnIleThrLeuGlyValGluAspGluLeuLys 257
 Db 732 ATGGTGTCAAGATCATCGTTCTCAATATATACCTGGGTGGTGAAGATGAATGAA 791
 Qy 258 HisAlaValGlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsn 277
 Db 792 TAGCAGATGTCATTTGGTGGCTGTAGTGTGCTTTAGGTGGTAAAGTTTCAA 851
 Qy 278 LeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsn 297
 Db 852 CAGTATAAGAGCGGAGTTTACACGACACTGAATGTGGCAGACTCCCATGGACGTAAC 911
 Qy 298 HisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLys 317
 Db 912 CATGCTGTTCTTGTGTTGGGTACGGTGTGAATGCGGTCCCTACTGCTCATTAAG 971
 Qy 318 AsnSerTrpGlyThrAsn 323
 Db 972 AACTCATGGGAGCAGAT 989
 RESULT 3
 LOCUS CK287181
 DEFINITION ES749903 Nicotiana benthamiana mixed tissue cDNA library,
 normalized, full-length Nicotiana benthamiana cDNA clone NEMB738 5',
 end, mRNA sequence.
 ACCESSION CK287181
 VERSION CK287181.1 GI:39863480
 KEYWORDS EST.
 SOURCE Nicotiana benthamiana
 ORGANISM Nicotiana benthamiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 (bases 1 to 985)
 AUTHORS Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A., Day,B.,
 Staskawicz,B., Jin,H. and Baker,B.
 TITLE Generation of EST sequences from Nicotiana benthamiana
 JOURNAL Unpublished (2003)
 COMMENT Other ESTs: ES749904
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr. Rockville, MD 20850, USA
 Email: potatoc-array@tigr.org
 Clones can be requested from TIGR via potatoc@tigr.org
 Seq primer: ATT TAG GTC ACA CTA TAG.
 Location/Qualifiers
 1. 985
 /organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NEMB738"
 /tissue_type="abiotic and biotic stress-treated leaves,
 callus tissue and root tissue"
 /lab_host="DH10B-Tona"
 /clone_lib="Nicotiana benthamiana mixed tissue cDNA
 library, normalized, full-length"
 /note="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: RNA was isolated from Nicotiana benthamiana
 tissues that include callus, roots from liquid culture

grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr), cold-stressed leaves (5 C 3 hr, 6hr), and pathogen challenged leaves (Pseudomonas syringae pv tomato 12 hr; Xanthomonas campestris pv campestris 12 hr, 18hr; Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas campestris pv vesicatoria 18hr). RNA was isolated from these tissues and pooled in approximately equal molar amounts."

ORIGIN

Alignment Scores:
 Pred. No.: 7,45e-113 Length: 985
 Score: 1039.00 Matches: 206
 Percent Similarity: 76.53% Conservative: 32
 Best Local Similarity: 66.24% Mismatches: 65
 Query Match: 55.66% Indels: 8
 DB: 14 Gaps: 5

US-10-087-714-2 (1-352) x CK287181 (1-985)

Qy 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlagly----- 22
 Db 52 CTCCTATTGCTCTTCGTCGCGGTGGCTTTTCGCGCGCACTTGCAGCGGCT 111
 Qy 23 ---PheGluGluAspAsnProIleArgSer---ValThrGlnArgProAspSerIleGlu 40
 Db 112 ACCTTCCGCGATGAGATCCGATCGACAAATTTGTTCTGACGGTTTACATGAGCTGGAG 171
 Qy 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
 Db 172 AAGCGAATTCCTCAAGTCGTCGCGCAAGACTCGTCGCTCTCTCTCTGCTGCTGCTTGT 231
 Qy 61 ArgArgTyrGlyIlySerTyrGlySerGluGluGluIleIlyIlyIlyIlyIlyIlyIly 80
 Db 232 CACAGGTATGGGAAGAGTACGAGACAGTGTGAGAGATTAAGCAAGGTTTCGAGGATTTC 291
 Qy 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgIlyAspLeuSerTyrThrLeuGly 100
 Db 292 TTGGACAATTTGAAGATGATTCGATCGATACAAAGAGGACTATCATACAACTCGGT 351
 Qy 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgIleGlyAlaAla 120
 Db 352 GTCAATGAGTTTACCAGACATAACATGGGAGTTCGAGAGTTCGAGAGAGGTTGGAGCAGCT 411
 Qy 121 GlnAsnCysSerAlaThrAlaHisGlyIlyAsnHisArgPheValAspGlyValLeuProVal 140
 Db 412 CAATACTGTTCCGCCACCACAAAGGCAATCTCAACTCACTCACTCACTGCTGCGGAG 471
 Qy 141 ThrArgAspTrpArgGluGlnGlyIleValSerProValIlyAspGlnGlySerCysGly 160
 Db 472 ACAGAAAGACTGGAGGGAAGCTGGGATTTGTCAGCCCACTCAAGAACCAAGGCGCAAGTGC 531
 Qy 161 Ser---TyrThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGly 179
 Db 532 TCTTGTGGACATTCAGCACTATCTGTTGCTAGAGCAGCATATGCGCAAGCATTTGGG 591
 Qy 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPhe 197
 Db 592 AAGCGAATTTCTCTATCTGACACGACCTTGTGGACTGTGCTGAGCTTTTAATACTTT 651
 Qy 198 GlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValIlyIlyIlyIlyIlyIly 216
 Db 652 GGCTGCAATGTTGGCTCCCATCACAGCCTTTGAGTATATTAATCAATGGTGTCTT 711
 Qy 217 AspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheIlyGlnGlu 236
 Db 712 GACACTGAAGAGCATATCCATACACCGGCAAGATGGCTTATGTAATTTCTCATCAGAA 771
 Qy 237 AsnValGlyValIlyValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeu 256
 Db 772 AATGTTGGTCAAGAGCATCGATCTTCTGCAATATTACCTGGTCTGAAGATGAATCA 831
 Qy 257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluValValIlyGlyPhe 276

Db 832 AATACGCGGTTGCATGGTTAGCCGCTTAGTAGCTTTTGGAGTGATAAAAGGTTTC 891
 Qy 277 AsnLeuTyrIlyValGlyValTyrSerSerAspThrCysGlyArgAspProMetAspVal 296
 Db 892 AAACAATACAGAGTGGTGTGTTACCAAGACACCAAGTGGGCATATCCCATGATGTA 951
 Qy 297 AsnHisAlaValLeuAlaValGlyTyrGlyVal 307
 Db 952 AACCATGCTGTCTTGTCTGTGGGTAGCGTGT 984

RESULT 4

CB341855

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Vitis vinifera

Vitis vinifera

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; Vitaceae; Vitis.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Mar 14, 2003 this sequence version replaced gi:28962822.

Contact: Douglas Cook, PhD

CBS Genome Facility

UC Davis, Plant Pathology

One Shields Ave, Davis, CA 95616, USA

Tel: 530 754 6561

Fax: 530 754 6617

Email: drcook@ucdavis.edu

Seq primer: ACGTACCGGACATATGCC.

Location/Qualifiers

1..829

/organism="Vitis vinifera"

/mol_type="mRNA"

/cultivar="Cabernet Sauvignon"

/db_xref="taxon:29760"

/clones="CA32EN002_IIIBF_H04"

/sex="Hermaphrodite"

/dev_stage="Mid-season leaf material, collected July 25, 2001."

/lab_host="DH5alpha"

/clone_lib="Cabernet Sauvignon Leaf - CA32EN"

/note="Organ: Leaf; Vector: pDNR; Site: 1: SfiI; Site: 2: SfiI; CA32EN is a cDNA library of Cabernet Sauvignon leaves. The leaves were collected on July 25, 2001, in Napa Valley, California, and represent leaves in mid-season development. These leaves were asymptomatic and verified to be non-infected with the bacterial pathogen, Xylella fastidiosa, based on a diagnostic assay using PCR and Xylella-specific primer pairs. cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:

5'-AAGCAGTGGTATCAACGAGTGGCCATACGCGCGG-3' and

5'-ATTCTAGCCCGGCGGCGACATG-DT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Alignment Scores:

Pred. No.: 1,31e-112 Length: 829

Score: 1035.00 Matches: 197

Percent Similarity: 79.71% Conservative: 23

Best Local Similarity: 71.38% Mismatches: 52

Query Match: 55.50% Indels: 4

```

DB: 14 Gaps: 3
US-10-087-714-2 (1-352) x CB341855 (1-829)
QY 42 AlalLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArg 61
Db 2 TCCGCTCCCGCTCGCTCGGACACCGGTACGCTACTCTCTTCGCGAGCTTCGCTCAC 61
QY 62 ArgTyrGlySerTyrGlySerGluGluGluLeuLysArgPheGlyLeuPheVal 81
Db 62 AGGTATGGAGAGATTACAAGCGGTGGACGAGATTAGCTGAGATTTCGATTTTCG 121
QY 82 GluAsnLeuAlaPheAlaArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyLe 101
Db 122 GAGAAATTTGAACTCATCAGATCCACCAAGAGAGGCTTGCCTTACTACTCTGCTGT 181
QY 102 AsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyValAlaAlaGln 121
Db 182 AATCAGTTTCGCTGATGACCTGGAGAGATTCCGCAGACACAGGTTGGAGCTGCTCAG 241
QY 122 AsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThr 141
Db 242 AACTGCTCTGCACCTTGAAGGGCAATCACAAGCTAACTGACGTATCTCTCTGAGACG 301
QY 142 ArgAspTrpArgGluGluGlyLeuValSerProValLysAspGlnGlySerCysGlySer 161
Db 302 AAAGATTGGAGAGAGATGGCATAGTCAAGCCAAATCAAGATCAAGTCACTGTGGATCT 361
QY 162 ---TrpThrPheSerThrThrGlyValAlaLeuGluAlaAlaLysThrGlnLeuThrGlySer 180
Db 362 TGCTGGACCTTTCAGCACCTGGAGCTCTAGAGGACGCTTACGCTCAGGCATTTGGGA 421
QY 181 -----ThrLeuSerGluGlnGluLeuValAspCysAlaSerAlaPheAsnAsnPheGly 199
Db 422 GGGATCTCTCTGCTGAGCAGCAGCTGTGGACTGTGCGGAGCTTCAATAAATTTTGA 481
QY 199 Cys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLeuAsp 217
Db 482 TGCCACGGGGATTGCCATCCCAAGCTTTTGAGTATACATAAATACATGTTGGCTTGAT 541
QY 218 ThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCysAsnPheLysGlnGluAsn 237
Db 542 ACTGAGGAAGCATATCTCTACACTGGACTAGATGGCACCTGCACAAATTTCTTCAGAAA 601
QY 238 ValGlyValLysValLeuAspSerIleAsnLeuThrLeuGlyAlaGluAspGluLeuLys 257
Db 602 ATTGGTGTCAAGTCTCGACTCTGTGAATATACCTGGGTGCTGAGATGATGATTAAG 661
QY 258 HisAlaValGlyLeuValArgProValSerValAlaPheGluValLysGlyPheAsn 277
Db 662 CATGCAGTTGCATTTGTTCTGCTCAGTGTGAGTGTGCTGAGTGTGCTGATTTCCGA 721
QY 278 LeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsn 297
Db 722 TTTTACAGAAAGAGATTACAAAGTGGAACTTGTGACAGCACTCCCATGGATGGAAC 781
QY 298 HisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
Db 782 CATGCTGTCTTCAGTCGGGTATGGAGTTGGATCTGCTGATGATATAC 829

RESULT 5
LOCUS CK295860 977 bp mRNA linear EST 15-DEC-2003
DEFINITION EST758574 Nicotiana benthamiana mixed tissue cDNA library,
normalized, full-length Nicotiana benthamiana cDNA clone NBMCX76 5'
end, mRNA sequence.
ACCESSION CK295860
VERSION CK295860.1 GI:39880659
KEYWORDS EST.
SOURCE Nicotiana benthamiana
ORGANISM Nicotiana benthamiana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

REFERENCE

1 (bases 1 to 977)
 Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
 Staskawicz, B., Jin, H. and Baker, B.

TITLE

Generation of EST sequences from *Nicotiana benthamiana*
 Unpublished (2003)

JOURNAL

Other ESTs: EST758575
 Contact: Robin Buell

COMMENT

The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org

FEATURES

Clones can be requested from TIGR via potato@tigr.org
 Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

Location/Qualifiers
 1..977

/organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NBMCX76"

/tissue_type="abiotic and biotic stress-treated leaves,
 callus tissue and root tissue"

/lab_host="DH10B-Tona"

/clone_lib="Nicotiana benthamiana mixed tissue cDNA
 library, normalized, full-length"

/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: RNA was isolated from *Nicotiana benthamiana*
 tissues that include callus, roots from liquid culture
 grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
 cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
 challenged leaves (*Pseudomonas syringae* pv tomato 12 hr;
Xanthomonas campestris pv campestris 12 hr, 18hr;
Pseudomonas syringae pv phaseolicola 18hr, and *Xanthomonas*
campestris pv vesicatoria 18hr). RNA was isolated from
 these tissues and pooled in approximately equal molar
 amounts."

ORIGIN

Alignment Scores:

Pred. No.: 2,21e-112 Length: 977
 Score: 1034.00 Matches: 208
 Percent Similarity: 76.51% Conservative: 33
 Best Local Similarity: 66.03% Mismatches: 65
 Query Match: 55.44% Indels: 9
 DB: 14 Gaps: 5

US-10-087-714-2 (1-352) x CK295860 (1-977)

QY 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGly----- 22
 Db 33 CTCCTATTGGCTCTCGTCGCGCGGTGGCTTTTCGCGCGCACTTGGCGACGGCT 92
 QY 23 ---PheGluGluAspAsnProIleArgSer---ValThrGlnArgProAspSerIleGlu 40
 Db 93 ACCTTGCGCGATGAGAAATCCGATCATGACAAATTTGTTCTGACGGTTTACATGAGTGGAG 152
 QY 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
 Db 153 AACGGAATTCCTCAAGTGTGCGCAAGACTGTCATGCTCTCTCTCTCTCTCTCTCTCTCT 212
 QY 61 ArgArgTyrGlyLysSerTyrGlySerGluGluGluLysLysArgPheGlyLysPhe 80
 Db 213 CACAGTATGGAGAGAGGTACGACAGCTTGGAGGAGATAAAGCAAGGTTCCGAGGTATTC 272
 QY 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGly 100
 Db 273 TTGGCAATTTGAAGATGATTCGATCGCATTAACAAAGAAAGGACTATCATCAAACTCGGT 332
 QY 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla 120
 Db 333 CTCATGAGTTTACCGACATTAACATGGGACGAGTTCCGAGACAGACAGGTTGGGACGCT 392
 QY 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140

DB:

14

Gaps:

3

US-10-087-714-2 (1-352) x CB341855 (1-829)

QY

42 AlalLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArg 61

Db

2 TCCGCTCCCGCTCGCTCGGACACCGGTACGCTACTCTCTTCGCGAGCTTCGCTCAC 61

QY

62 ArgTyrGlySerTyrGlySerGluGluGluLeuLysArgPheGlyLeuPheVal 81

Db

62 AGGTATGGAGAGATTACAAGCGGTGGACGAGATTAGCTGAGATTTCGATTTTCG 121

QY

82 GluAsnLeuAlaPheAlaArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyLe 101

Db

122 GAGAAATTTGAACTCATCAGATCCACCAAGAGAGGCTTGCCTTACTACTCTGCTGT 181

QY

102 AsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyValAlaAlaGln 121

Db

182 AATCAGTTTCGCTGATGACCTGGAGAGATTCCGCAGACACAGGTTGGAGCTGCTCAG 241

QY

122 AsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThr 141

Db

242 AACTGCTCTGCACCTTGAAGGGCAATCACAAGCTAACTGACGTATCTCTCTGAGACG 301

QY

142 ArgAspTrpArgGluGluGlyLeuValSerProValLysAspGlnGlySerCysGlySer 161

Db

302 AAAGATTGGAGAGAGATGGCATAGTCAAGCCAAATCAAGATCAAGTCACTGTGGATCT 361

QY

162 ---TrpThrPheSerThrThrGlyValAlaLeuGluAlaAlaLysThrGlnLeuThrGlySer 180

Db

362 TGCTGGACCTTTCAGCACCTGGAGCTCTAGAGGACGCTTACGCTCAGGCATTTGGGA 421

QY

181 -----ThrLeuSerGluGlnGluLeuValAspCysAlaSerAlaPheAsnAsnPheGly 199

Db

422 GGGATCTCTCTGCTGAGCAGCAGCTGTGGACTGTGCGGAGCTTCAATAAATTTTGA 481

QY

199 Cys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLeuAsp 217

Db

482 TGCCACGGGGATTGCCATCCCAAGCTTTTGAGTATACATAAATACATGTTGGCTTGAT 541

QY

218 ThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCysAsnPheLysGlnGluAsn 237

Db

542 ACTGAGGAAGCATATCTCTACACTGGACTAGATGGCACCTGCACAAATTTCTTCAGAAA 601

QY

238 ValGlyValLysValLeuAspSerIleAsnLeuThrLeuGlyAlaGluAspGluLeuLys 257

Db

602 ATTGGTGTCAAGTCTCGACTCTGTGAATATACCTGGGTGCTGAGATGATGATTAAG 661

QY

258 HisAlaValGlyLeuValArgProValSerValAlaPheGluValLysGlyPheAsn 277

Db

662 CATGCAGTTGCATTTGTTCTGCTCAGTGTGAGTGTGCTGAGTGTGCTGATTTCCGA 721

QY

278 LeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsn 297

Db

722 TTTTACAGAAAGAGATTACAAAGTGGAACTTGTGACAGCACTCCCATGGATGGAAC 781

QY

298 HisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313

Db

782 CATGCTGTCTTCAGTCGGGTATGGAGTTGGATCTGCTGATGATATAC 829

RESULT 5

LOCUS

CK295860

DEFINITION

EST758574 Nicotiana benthamiana mixed tissue cDNA library,
 normalized, full-length Nicotiana benthamiana cDNA clone NBMCX76 5'
 end, mRNA sequence.

ACCESSION

CK295860

VERSION

CK295860.1

KEYWORDS

EST.

SOURCE

Nicotiana benthamiana

ORGANISM

Nicotiana benthamiana

Eukaryota; Viridiplantae;

Streptophyta; Embryophyta;

Tracheophyta;

Spermatophyta; Magnoliophyta;

eudicotyledons; core eudicots;

```

Db      393 CAAAACTGTTTCAGCCACCAAGGGCAATCTCAAACTCACTAACTGCTGCGCGAG 452
QY      141 ThrArgaspTtArgGluGlnGlylleValSerProVallyAspGlnGlySerCysGly 160
Db      453 ACAGAAAGCTGGAGGAGCTGGGATGTCAGCCCACTCAAGAACAGGCGCAAGTGCAG 512
QY      161 Ser---TtrThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGly 179
Db      513 TCTTGCTGGACATTCAGCACTACTCTGTCAGTACAGGCTAGAGACGATATGGCCCAAGCATTTGGG 572
QY      180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnPhe 197
Db      573 AAGGGAATTTCTCTATCTGACACAGCCTTGTGGATGTCGTGGAGCTTTTAAATACATT 632
QY      198 GlyCys----GlyGlyLeuProSerGlnAlaPheGluTyrVallySerPyrAsnGlyGlylle 216
Db      633 GCGTGCATGTTGGTGGCTCCCATCACAAGCCTTTGAGTATATTAATCCATGCTGCTT 692
QY      217 AspThrGluGlnThrTyrProTyrLeuGlyValMetGlylleCysAsnPhelysGlnGlu 236
Db      693 GACACTGAAGACGATATCCATACACCGCGCAAGATGGCTTATGTAATTCATCAGAA 752
QY      237 AsnValGlyVallyValleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeu 256
Db      753 AATGTTGTCGTCACAGTATGATGCTGCTCAATATTAACCTGGGTGCTGAAGATGAATA 812
QY      257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVallyGlyPhe 276
Db      813 AAATACGCGGTTGCAATGTTAGCGCGCTTATGATAGCTTTTCAGGTGATAAAGTTTC 872
QY      277 -AsnLeuTyrIlyAspGlyValTyrSerSerAspThrCysGlyArgAspProMetAspVa 296
Db      873 AAAAATAATACAGAGTGGTGTGTACACACGACCGAATGTGGCATACTCCCATGGATGT 932
QY      296 LasnHisAlaValLeuAlaValGlyTyrGlyValGluAspGly 310
Db      933 AAACCATGCTGTTCTTCTGTGGGTAGCGGTGTTGAATGTT 975

RESULT 6
LOCUS   CF212812
DEFINITION   CGF1000651_H04 Vitis vinifera cv. cabernet sauvignon Stem - CAST
            Vitis vinifera cDNA clone CAST0005_IF_H04 5', mRNA sequence.
ACCESSION   CF212812
VERSION     CF212812.1 GI:33407185.
KEYWORDS   EST.
SOURCE     Vitis vinifera
ORGANISM   Vitis vinifera
REFERENCE  1 (bases 1 to 1005)
AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook,D.
TITLE      Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon' berries at various developmental stages
JOURNAL    Unpublished (2003)
COMMENT    Contact: Douglas Cook, PhD
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcoc@ucdavis.edu
Seq primer: ACGGTACCGGACATATGCC.
            Location/Qualifiers
            1..1005
            /organism="Vitis vinifera"
            /mol_type="mRNA"
            /cultiVar="Cabernet Sauvignon"
            /db_xref="taxon:29760"
            /clone="CAST0005_IF_H04"
            /sex="Hermaphrodite"

```

```

/dev_stage="Pre-bloom (10-11 days before bloom)"
/lab_host="DH5alpha"
/clone_lib="Vitis vinifera cv. cabernet sauvignon Stem - CAST"
/notes="Organ: Stem; Vector: pDNR; Site 1: sf11; Site 2: sf11; CAST is a cDNA library of Vitis vinifera cv. 'Cabernet Sauvignon' stems. Samples were collected May 13, 2002 from pre-bloom plants (10-11 days before bloom), pre-veraison. Sampled vines were located at the University of California, Davis. Experimental vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAACGAGTGGCCATACGCCGGG-3', and
5'-ATTCTAGAGCGGCGGCGGCGCATG-3' (30'NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

```

ORIGIN

```

Alignment Scores:
Pred. No.:      5 48e-111      Length:      1005
Score:          1022.50      Matches:      202
Percent Similarity: 74.68%      Conservatives: 31
Best Local Similarity: 64.74%      Mismatches: 61
Query Match:      54.83%      Indels:      18
DB:              14          Gaps:        4

US-10-087-714-2 (1-352) x CF212812 (1-1005)
QY      10 LeuPheLeuValSerAlaLeuSerValAlaLeu----- 20
Db      72 CTCTCGTGTGCGCGCGCTCTGATCTGCTCTGCGCGCTGCTTCCGGGAGGCTGAC 131
QY      21 -----AlaGlyPheGluGluAspAsnProIleArgSerValThrGlnA-gPro 36
Db      132 CATCATTTCCGATCAAGCTTCGATGAAGAGAACCAATAGACTGTATCGGACAGCAT 191
QY      37 AspSerIleGluProAlaIleGluGlyValLeuGlySerCysArgHisAlaPheHisPhe 56
Db      192 CCGGACTGGATGCTCGCTCGGCTGATCGGGACACGCGTCTACTCCTTCTCCTTCT 251
QY      57 AlaArgPheAlaArgArgTyrGlySerTyrGlySerGluGluGluIleLysLysArg 76
Db      252 GCGAGCTTGTCTCAGAGTATGGGAAGAGTTACAGACGCGGTGGACGAGATTAAGCTGAGA 311
QY      77 PheGlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSer 96
Db      312 TTGAGATTTTCTCGGAGATTTGAACTCATCATGATCCACACACAGAAAGGCTTGCCT 371
QY      97 TyrThrLeuGlylleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArg 116
Db      372 TATATCTTAGCTGTAAATCAGTTTCGCTGATTTGGACCTGGGAAGAGTTCCGACACACAGG 431
QY      117 LeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGly 136
Db      432 TTGGGAGCTGCTCAGAACTGCTCTGCCACCTTGAAGGGCAATCACAAAGCTAACTGACGTT 491
QY      137 ValLeuProValThrArgAspTyrArgGluGlnGlnIleValSerProVallyAspGln 156
Db      492 ATCTCTCTCGAGACAAAGATTTGAGAGAGATGGCATAGTCAGCCCAATCAAGATCAA 551
QY      157 GlySerCysGlySer---TtrThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThr 175
Db      552 GGTCACTGTGGATCTTGTGCTGACTTTGAGCACCCTGAGAGCTCTAGAGGCGAGCTTACGCT 611
QY      176 GlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAla 193
Db      612 CAGGCAATTTGGAGGGGATCTCTCTGCTGAGGCGAGCTTTGGAGCTTGTGCGGAGCT 671
QY      194 PheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVallySer 212
Db      672 TTCAATAACTTTGGATGCGACCGGGGATGGCATCTCCCAAGCTTTTGTAGTACATCAATAAC 731
QY      213 AsnGlyGlylleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlylleCysAsn 232

```



```

Db      848  ACCGAGTTTACACGAGCACTGAATGTGGGACACTCCCATCGACGGTAACCATGCTGTT 907
Qy      301  LeuAlaValGlyTyrGlyVal 307
Db      908  CTTGCTGTGGGCTACCGTGT 928

RESULT 8
CK257583      920 bp mRNA linear EST 12-DEC-2003
LOCUS      EST741220 potato callus cDNA library, normalized and full-length
DEFINITION      Solanum tuberosum cDNA clone POC649 5' end, mRNA sequence.
ACCESSION      CK257583
VERSION
KEYWORDS
SOURCE
ORGANISM      Solanum tuberosum (potato)

REFERENCE
AUTHORS      Buell, C.R., Hart, A., Zismann, V., Karameycheva, S.A. and Baker, B.
TITLE      Generation of ESTs from potato callus tissue
JOURNAL      Unpublished (2003)
COMMENT      The Institute for Genomic Research
              9712 Medical Center Dr, Rockville, MD 20850, USA
              Email: potato-array@tigr.org
              Clones can be requested from TIGR via potato@tigr.org
              Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES
source
location/Qualifiers
1..920
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POC649"
/tissue_type="callus"
/lab_host="DH10B-TorA"
/clone_lib="potato callus cDNA library, normalized and
full-length"
/notes="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Solanum tuberosum var.
Kennebec callus tissue grown on solid media."

ORIGIN
Alignment Scores:
Pred. NO.:      6.8e-109      Length:      920
Score:      1004.50      Matches:      204
Percent Similarity:      76.97%      Conservative:      30
Best Local Similarity:      67.11%      Mismatches:      61
Query Match:      53.86%      Indels:      9
DB:      14      Gaps:      6

US-10-087-714-2 (1-352) x CK257583 (1-920)
Qy      9  LeuLeuPheLeuValSerAlaLeu---SerValAlaLeuAlaGly-----PheGlu 24
Db      11  CTATGGTCTTGTTCGGCGCTTTTCGCTCTGCTGCTGCTGCGACCGGACCTTCGCC 70
Qy      25  GluAspAsnProIleArgSerValThrGlnArgProAspSerIleGluProAlaIleLeu 44
Db      71  GATGAGAATCCGATCAGCAAGTAGTAGTT---ACCGAAGAGCTGGAGAACGGAATTCCT 127
Qy      45  GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgTyrGly 64
Db      128  CAGTCGTGGCGCAGACTCGGAATGCTCTCTTCGCTCGCTTTCGCTATCAGGCACTGG 187
Qy      65  LysSerTyrGlySerGluGluIleLysArgPheGlyIlePheValGluAsnLeu 84
Db      188  AAAAGGTACGAGTCGCTGAGGAGATCAAGCAAGAGGTTTCGAGATATTTTGGCAATCTG 247
Qy      85  AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104

```

```

Db      248  AAGATGATCCGATCGCATAAACAGCAAGGACTATCATACAACTCGGTGTCATGAGTTT 307
Qy      105  AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
Db      308  ACCGACCTAACATGGGATGAGTTCGCTAGACACAAGATTGGGGGATCTCAAAAGTGTCT 367
Qy      125  AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
Db      368  GCCACTACAAAGGCAATCTCAAGCTAACTAAGCTGTTCTGCCAGAGAGAGGACTGG 427
Qy      145  ArgGluGlnGlyIleValSerProValIleAspGlnGlySerCysGlySer---TrpThr 163
Db      428  AGGGAAGATGGTATTGTTAGCCAGTGAAGGCACAGGGCAAGTGGGATCTTGCTGGACA 487
Qy      164  PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
Db      488  TTCAGCACTACTGTGTCACCTAGAGGCAGCATATGCCCAAGCATTTGGGAAGGGGAATCTCT 547
Qy      192  LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
Db      548  CTGTCAGAGCAGCAGCTTGTGGACTGTGCTGGAGCTTTTAATACTTTGGCTGCAATGGG 607
Qy      201  GlyLeuProSerGlnAlaPheGluTyrValIleGlyTyrAsnGlyGlyIleAspThrGluGln 220
Db      608  GGGTGGCATCACAAAGCCTTTGAGTACATTAAATTCATGCTGTGCTTGACACACTGAAGA 667
Qy      221  ThrTyrProTyrLeuGlyValIleGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
Db      668  GCATATCCATACACCGGCAAGATGGCATATGTAATTCACAAAGCAAAATATTGGTGTC 727
Qy      241  LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
Db      728  AAGTCACTACGTTCCTCATATTAATCCCTGGTGTGCTGAAGATGAATCAATATCGCAGTT 787
Qy      261  GlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
Db      788  GCATGGTGTAGCCCTGTAGTGTTCCTTTGAGGTGTAAAGGTTTCAACACAGTATTAAG 847
Qy      281  LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
Db      848  ACCGAGGTTTACACCGACACTGAATGTGGCAGACTCCCATCGGACGTAACCATGCTGTT 907
Qy      301  LeuAlaValGly 304
Db      908  CTTGCTGTGGGC 919

RESULT 9
CK257583
LOCUS
DEFINITION
ACCESSION      CK257583
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

748 bp mRNA linear EST 01-MAY-2003
 CB982399 CAB70006_IVaF_A04 Cabernet Sauvignon Berry Post-Veraison - CAB7
 Vitis vinifera cDNA clone CAB70006_IVaF_A04 5', mRNA sequence.

CB982399.1 GI:30305605

EST.

Vitis vinifera

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; Vitaceae; Vitis.

1 (bases 1 to 748)

Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and

Cook, D.

Expressed sequence tags from cabernet sauvignon berries at various

developmental stages

Unpublished (2003)

Contact: Douglas Cook, PhD

CAES Genome Facility

UC Davis, Plant Pathology

One Shields Ave, Davis, CA 95616, USA

Tel: 530 754 6561

Fax: 530 754 6617

Email: drcook@ucdavis.edu

Seq primer: ACGGTACCGGACATATGTC.

FEATURES		Location/Qualifiers	
source		1..748	
		/organism="Vitis vinifera"	
		/mol_type="mRNA"	
		/cultivar="Cabernet Sauvignon"	
		/db_xref="taxon:29760"	
		/clone="CAB70006 IVaF A04"	
		/sex="Hermaphrodite"	
		/dev_stage="Post-Veraison, 18-19 brix"	
		/lab_host="DHSalpa"	
		/clone_lib="Cabernet Sauvignon Berry Post-Veraison - CAB7"	
		/note="Organ: Berry; Vector: pDNR; Site 1: SfiI; Site 2: SfiI; CAB7 is a cDNA library of Cabernet Sauvignon Clone 8 berries. Samples were collected post-veraison from field-grown vines during stage III of berry growth at 89 days after full bloom. Berries soluble solid concentration ranged between 18-19 brix. Sampled vines were located at the University of California, Davis, Experimental Vineyard. cDNAs were made by oligo-dT priming and directly cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGTATCAACGAGAGTGGCCATTACGCGCGG-3' and 5'-ATTCTAGGCGGCGGCGGAGCATG-dt(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."	
ORIGIN			
Alignment Scores:		4.67e-107	Length: 748
Pred. No.:		988.00	Matches: 187
Score:		81.93%	Conservative: 17
Percent Similarity:		75.10%	Mismatches: 41
Best Local Similarity:		75.98%	Indels: 4
Query Match:		14	Gaps: 3
DB:			
US-10-087-714-2 (1-352) x CB982399 (1-748)			
Qy	72	GlutLeuLysArgPheGlyLeuValGluAsnLeuAlaPheIleArgSerThrAsn	91
Db	2	GAGATTAAAGCTGAGATTCGAGATTTCGAGAGAAATTGAACTCATCAGATCCCAAC	61
Qy	92	ArgLysAspLeuSerThrLeuGlyLeuAsnGlnPheAlaAspLeuThrTrpGluGlu	111
Db	62	AGAAAGGCGTTCCTTATCTAGCTGTTAATCAGTTCGCTGATTCGACCTGGGAAGAG	121
Qy	112	PheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHis	131
Db	122	TTCCGCAGACACAGGTGGGAGCTGCTCAGAACATGCTCTGCACCTTGAAGGGCAATCAC	181
Qy	132	ArgPheValAspGlyValLeuProValThrArgAspTrpArgGluGlnGlyValSer	151
Db	182	AAGCTAACTGAGCTTATCTTCCTCAGACAGAAAGATTGGAGAGAAGATTCATATTCAGC	241
Qy	152	ProValLysAspGlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeu	170
Db	242	CCAATCAAGATCAAGGTCACTGTGGATCTTGTCTGAGCTTTTCAGCACCACTTGGAGCTTA	301
Qy	171	GluAlaAlaThrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuVal	188
Db	302	GAGGAGCTTACGCTCAGCGATTGGGAGGGGATCTCTCTCTGAGCAGCAGCTGTG	361
Qy	189	AspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPhe	207
Db	362	GACTGTGCGGAGCTTTCAATAACTTTTGATGTCACCGGGGATTGCCATCCCAAGCTTTT	421
Qy	208	GlnThrValLysThrAsnGlyLeuAspThrGluGlnThrThrTrpProThrLeuGlyVal	227
Db	422	GAGTACATCAATAACAAAGGTGGCGCTTGATCTAGGGAAGCATATCTCTTACCTGGACTA	481
Qy	228	MetGlyIleCysAsnPheLysGlnGluAsnValGlyValLysValIleAspSerIleAsn	247
Db	482	GATGGCACTGCATAATTTCTTCAGAAAATATTGGTGTTCAGTCTTCGACTCTGTGAAT	541
Qy	248	IleThrLeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSer	267
Db	542	ATTACCTGGTGGTGAAGATGAGTAAAGCATGTCAGTTGCATTGCTTCAGTCAAGT	601
Qy	268	ValAlaPheGluValValLysGlyPheAsnLeuThrLysGlyValThrSerSerAsp	287
Db	602	GTGGCATTTTGAAGTGGTCCATGATTTCGATTTTACAGAAAGAGGAGTTTACACAAAGTGA	661
Qy	288	ThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyThrGlyVal	307
Db	662	ACTTGTGGCAGCACTCCCATGATGATGAACCATGCTGTTCTTGCACTGGGTATGAGTT	721
Qy	308	GluAspGlyLeuProThrTrpLeuLeu 316	
Db	722	GAGATGGTGTAGCATACTGGTCTATC 748	
RESULT 10			
BG594735			
LOCUS			
DEFINITION		814 bp mRNA linear EST 07-MAR-2003	
		EST493413 cSTS Solanum tuberosum cDNA clone cSIS8E24 5' sequence,	
		mRNA sequence.	
ACCESSION		BG594735	
VERSION		BG594735.1 GI:13612875	
KEYWORDS		EST.	
SOURCE		Solanum tuberosum (potato)	
ORGANISM		Solanum tuberosum	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
		asterids; lamids; Solanales; Solanaceae; Solanum.	
AUTHORS		van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A.,	
		Bougri, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.	
TITLE		Generations of ESTs from sprouting potato eyes	
JOURNAL		Unpublished (2000)	
COMMENT		Contact: Robin Buell	
		The Institute for Genomic Research	
		9712 Medical Center Dr, Rockville, MD 20850, USA	
		Email: potato-array@igr.org	
		This clone can be obtained from the University of Arizona Genomics	
		Institute. Orders can be made through URL:	
		http://genome.arizona.edu/orders/	
Seq primer:		M13F-R.	
FEATURES		Location/Qualifiers	
source		1..814	
		/organism="Solanum tuberosum"	
		/mol_type="mRNA"	
		/cultivar="Kennebec"	
		/db_xref="taxon:4113"	
		/clone="cSTS8E24"	
		/tissue_type="sprouting eyes: from tubers"	
		/dev_stage="12-14 weeks post harvest"	
		/lab_host="SOLR"	
		/clone_lib="cSTS"	
		/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."	
ORIGIN			
Alignment Scores:		6.99e-107	Length: 814
Pred. No.:		987.00	Matches: 193
Score:		80.67%	Conservative: 24
Percent Similarity:		71.75%	Mismatches: 48
Best Local Similarity:		52.92%	Indels: 5
Query Match:		12	Gaps: 3
DB:			
US-10-087-714-2 (1-352) x BG594735 (1-814)			
Qy	56	PheAlaArgPheAlaArgArgThrGlyLysSerThrGlySerGluGluLeuLysLys	75
Db	7	TTGCTCTGCTTTCCTATCATCGGCATCGGAAAGTACGAGTCCGTTGAGGAGATCAAGCAA	66

QY 76 ArgPheGlyLeuPheValGluAlaGlnLeuAlaPheIleArgSerThrAsnArgLysAspLeu 95
 Db 67 AGTTTCAGATATTTTGGCAATCTTGAAGATGATCCGATCGCATACGATACGAAAGACTA 126
 QY 96 SerTyrThrLeuGlyLeuAlaGlnLeuPheAlaAspLeuThrTrpGluGluPheArgThrAsn 115
 Db 127 TCATACAAATCGGTCTCAATGATGTTTACCGACCTAACATGGATGGATGGATTCGGTAGACAC 186
 QY 116 ArgLeuGlyAlaAlaGlnLeuPheValAlaGlnLeuPheAlaHisGlyAsnHisArgPheValAsp 135
 Db 187 AAGTTGGGGCATCTCAGAACTCTTCTGACACTACAAAGGGCAATCTCAAGTAACACTAAC 246
 QY 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAsp 155
 Db 247 GTCGTTCTGCGACAGACGAGGACTGCGAGGAAGATGTTATTTAGCCAGTGAAGGCA 306
 QY 156 GlnGlySerCysGlySer---TrpThrPheSerThrGlyValAlaGluAlaAlaTyr 174
 Db 307 CAGGCAAGTGGCGATCTTCTGTCACATTCAGCACTACTGGTGCCTAGGCGACATAT 366
 QY 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
 Db 367 GCCCAAGCATTTGGGAAGGAATCTCTCTGTCAGACGACGAGCTTGTGGACTGTGCTGA 426
 QY 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
 Db 427 GCTTTTAAATAACTTTGGCTGCAATGGGGGTTGCCATCAAGCCCTTTCAGTACATTTAAA 486
 QY 212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCys 231
 Db 487 TTCAATGGTGGTCTTGACACTGAAGAAGCATATCCATACCGGCAAGAATGGCATATGT 546
 QY 232 AsnPheLysGlnGluAlaGlnValGlyValLysValIleAspSerIleAsnIleThrLeuGly 251
 Db 547 AAATCTCTCAAGCAAAATATTTGGTGTCAAAGTCATCATCTCTCAATATTACCCCTGGGT 606
 QY 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 Db 607 GCTGAAGATGAATGAAATACGAGTTCGATTTGGTTAGCCCTGTAGTGTCTCTTTGAG 666
 QY 272 ValValLysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArg 291
 Db 667 GTGGTAAAGGTTTCAACACAGTATAAGACGGAGTTTACACGACGACTGAATGTGGCGAC 726
 QY 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311
 Db 727 ACTCCATGACGATTAACCATGCTGCTCTGTCTGGGTACGGTGGTGAATAATGGCGTT 786
 QY 312 ProTyrTrpLeuLysAsnSerTrp 320
 Db 787 CCCTACTGGTTCAT-AAGAACTCATGG 812
 RESULT 11
 CK294622
 LOCUS CK294622 965 bp mRNA linear EST 15-DEC-2003
 DEFINITION EST757336 Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length Nicotiana benthamiana cDNA clone NEMCO76 5', end, mRNA sequence.
 ACCESSION CK294622
 VERSION CK294622.1 GI:39878196
 KEYWORDS EST.
 SOURCE Nicotiana benthamiana
 ORGANISM Nicotiana benthamiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 (bases 1 to 965)
 AUTHORS Buell, C.R., Hart, A., Zismann, V., Karanycheva, S.A., Day, B., Skaskawicz, B., Jin, H. and Baker, B.
 TITLE Generation of EST sequences from Nicotiana benthamiana
 JOURNAL Unpublished (2003)
 COMMENT Contact: Robin Buell

The Institute for Genomic Research
 9712 Medical Center Dr. Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from TIGR via potato@tigr.org
 Seq primer: AT1 TAG GIG ACA CTA TAG.
 Location/Qualifiers
 1..965
 /organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NEMCO76"
 /tissue_type="abiotic and biotic stress-treated leaves, callus tissue and root tissue"
 /lab_host="DH10B-Tona"
 /library="Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length"
 /note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI; supplier: RNA was isolated from Nicotiana benthamiana tissues that include callus, roots from liquid culture grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr), cold-stressed leaves (5 C 3 hr, 6hr), and pathogen challenged leaves (Pseudomonas syringae pv tomato 12 hr; Xanthomonas campestris pv campestris 12 hr, 18hr; Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas campestris pv vesicatoria 18hr). RNA was isolated from these tissues and pooled in approximately equal molar amounts."
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.72e-106 Length: 965
 Score: 981.00 Matches: 195
 Percent Similarity: 75.67% Conservative: 32
 Best Local Similarity: 65.00% Mismatches: 65
 Query Match: 52.60% Indels: 8
 DB: 14 Gaps: 5
 US-10-087-714-2 (1-352) x CK294622 (1-965)
 QY 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGly----- 22
 Db 64 CTCCTATTGGCTCTCGTCGTCGCGCGGTGGCGCTTTTCGCCCGCCGACCTTGGCGACCGGCT 123
 QY 23 ---PheGluGluAspAsnProIleArgSer---ValThrGlnArgProAspSerIleGlu 40
 Db 124 ACCTTCCCGATGAGATCCGATCAGACAAATTTCTTCACGGTTCATACATGACGTGGAG 183
 QY 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
 Db 184 AACGGAATTTCTCAAGTCGTCGCGCAAGACTCGTCATGCTCTCTCTTCGCTCGCTTTTGT 243
 QY 61 ArgArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPheGlyIlePhe 80
 Db 244 CACAGGTATGGAGAGGTACGACAGACAGTTCGAGAGATAAAGCAAGGTTTCGAGGTATTC 303
 QY 81 ValGluLeuLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGly 100
 Db 304 TTGGACAAATTTGAAGATGATTCGATCGCATACAAAGAAAGGACTATCATCAAACTCGGT 363
 QY 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLysValAlaAla 120
 Db 364 GTCAATGAGTTTACCGACATAACATGGAGCAGTTCGCGAGACAGACAGTTCGGAGCAGCT 423
 QY 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140
 Db 424 CAAAACTGTTTCAGCCACCAAGGCAATCTCAAACTCACTACCTGTGTCTCTCCCGGAG 483
 QY 141 ThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGly 160
 Db 484 ACGAAGACTGGAGGAGGCTGGGATTTGTAGCCCACTCAAGAACCAAGGCAAGTGGCGGA 543
 QY 161 Ser---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGly 179

The Institute for Genomic Research
 9712 Medical Center Dr. Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from TIGR via potato@tigr.org
 Seq primer: AT1 TAG GIG ACA CTA TAG.

Db 544 TCTTCTGGACATTCAGCACTACTGGTGCACCTAGAGCAGCATATGGCCAAAGCATTTTGGG 603
 QY 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAspPhe 197
 Db 604 AAGGAAATTTCTCTATCTATGAACAGCAGCTTGGAGCTGTGGAGCTTTTAATAACTTT 663
 QY 198 GlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLe 216
 Db 664 GCTTCGAATGGTGGCTCCCATCAACAGCCTTTGAGTATATATAATCCAATGGTGTCTT 723
 QY 217 AspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuValMetGlyLeuValMetGlyLeuGlu 236
 Db 724 GACACTGAAGAGCATATCCATACACCGGCAAGATGGCTTATGTAATTTCTCATCAGAA 783
 QY 237 AsnValGlyValLysValLysSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 256
 Db 784 AATGTGTGTGTCAAGAGTCATGCTCTCTCAATATATACCTGGGTGTGTAAGATGAACATA 843
 QY 257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluValValLysGlyPhe 276
 Db 844 AATACGGGGTTCGATGGTGGCCCTTTAGTATAGCTTTTGGAGTGATATAAAGGTTTC 903
 QY 277 AsnLeuTyrLysGlyValLysGlyValLysGlyValLysGlyValLysGlyValLysGlyVal 296
 Db 904 AAACAATACAGAGTGGTGTATACACGAGCACCAATGTGGCCATCTCCCATGATGTA 963
 RESULT 12
 LOCUS BQ281381
 DEFINITION BQ281381.1 GI:20548869
 ACCESSION BQ281381
 VERSION EST
 KEYWORDS
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 731)
 Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
 Han,P.S., Heia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen,H.T.,
 Rauch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
 The structure and function of the expressed portion of the wheat
 genomes - Normalized shoot cDNA library
 Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: candersn@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer.
 Location/Qualifiers
 1..731
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /culti_var="Chinese Spring"
 /db_xref="taxon:4585"
 /clone="WHE3020_F06_L12"
 /tissue_type="Etiolated shoot"
 /dev_stage="Five day old seedling"
 /lab_host="E. coli DH10B"
 /clone_lib="Wheat unstressed seedling shoot normalized
 cDNA library"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid
 pBluescript SK, Site 1: EcoRI; Site 2: XhoI; Seeds were
 surface-sterilized, germinated and grown aseptically in
 the dark at room temperature on filter paper with water,

nystatin and cefotaxime in covered crystallization
 dishes. Roots were harvested. The tissue, total RNA, and
 poly(A) RNA were prepared, a cDNA library was made in the
 10 Close lab (Choi, Close, Fenton) at the University of
 California, Riverside. The cDNA clones were in vivo
 excised to give pBluescript phagemids before
 normalization was carried out. The mass excision of
 phagemid library and normalization were done in HT Nguyen
 lab by D. Zhang at Texas Tech University. Normalization
 protocol used was that of Soares'. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors).

ORIGIN

Alignment Scores: 1,41e-105 Length: 731
 Pred. No.: 975.50 Matches: 182
 Score: 86.36% Conservative: 27
 Percent Similarity: 75.21% Mismatches: 28
 Best Local Similarity: 52.31% Indels: 5
 Query Match: 13 Gaps: 4
 DB:

US-10-087-714-2 (1-352) x BQ281381 (1-731)

QY 78 GlyLeuPheValGluAsnLeuAlaPheLeuArgSerThrAsnArgLysAspLeuSerTyr 97
 Db 4 GGGATCTTCTCCGAGAGCCTCGAGCTGCTCCACCAACCGAGGGGCGCTCCCTAC 63
 QY 98 ThrLeuGlyLeuAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
 Db 64 CGCTCGGCATCAACCGTTTCGCGACATGAGCTGGAGGAGTTCCAGCGAGCGGCTC 123
 QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAsp---Gly 136
 Db 124 GCGCGGCGCAGACTGCTCGCGACGCTCGCGGCAACCAACCGATGCGCGACCGCC 183
 QY 137 ValLeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGln 156
 Db 184 GTCCTCCGCGACCAAGACTGGAGGAGGATCGTGAGCCCGTGGAGCCCGTGAAGACCA 243
 QY 157 GlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThr 175
 Db 244 GGTCACTGTGTCTCTGTGGACCTTCAGCACCACTGTTCTCTTCTTCTTCTTCTTCT 303
 QY 176 GlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAla 193
 Db 304 CAGGCACCTGGAAACCTGTCTCTCTTCTGAGCAGCAGCTGGTGTGTGTGTGTGTG 363
 QY 194 PheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyr 212
 Db 364 TACAATAATTTCCGATGCAGTGGAGGCTTACCATCTCAGGCTTTTGTACATCAATAC 423
 QY 213 AsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCysAsn 232
 Db 424 AATGTGTGCTTGCACACGAGAGGCTTACCTTACACGGGTGTCAATGGCATCTGTCT 483
 QY 233 PheLysGlnGluAsnValGlyValLysValLysSerLeuAsnLeuThrLeuGlyAla 252
 Db 484 TACAAGCTGAAAGAGTTGGAGTCAAGTTTGGACTCCGTTCATCATCACCTGGTCT 543
 QY 253 GluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluVal 272
 Db 544 GAGGATGAGCTGAAGAAATGCTGTGTGGAGTTGCTTCTGCTGCTTGTGTGTGTGTG 603
 QY 273 ValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAsp 292
 Db 604 ATCAACGTTTCAGAGGTACAGAGTGGAGTTTACACAGTACCATTGTGGAACTTCT 663
 QY 293 ProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLeuPro 312
 Db 664 CCAATGATGTGAACCAACCGCTGTCTGCGCCGTGGCTATGCTGCGAAATGGGTTCCC 723
 QY 313 TyrTrp 314

```

Db      724 TACTGG 729
|||||
RESULT 13
CB892658
LOCUS   768 bp mRNA linear EST 24-APR-2003
DEFINITION HSG6450 HOGA Medicago truncatula cDNA clone HOGA-19K6, mRNA
sequence.
ACCESSION CB892658
VERSION   1 GI:30099826
KEYWORDS EST.
SOURCE   Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids 1; Fabiales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 768)
AUTHORS Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
Utterback,T., Cho,J. and Fraser,C.M.
TITLE   ESTs from roots of Medicago truncatula treated with
oligogalacturonides of DP 6-20
JOURNAL Unpublished (2001)
COMMENT Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahn@ccrc.uga.edu
TIGR sequence name: WTMCV63TK
More information is available at: www.medicago.org
Seq primer: SKned (CTA GAA CTA gtg gat CC).
FEATURES
source
1..768
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="HOGA-19K6"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/clone_lib="HOGA"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
ORIGIN
Alignment Scores:
Pred. No.: 6.9e-105 Length: 768
Score: 970.00 Matches: 181
Percent Similarity: 81.25% Conservative: 27
Best Local Similarity: 70.70% Mismatches: 44
Query Match: 52.01% Indels: 4
DB: 14 Gaps: 3
US-10-087-714-2 (1-352) x CB892658 (1-768)
QY 65 LysSerTyrGlySerGluGluGluLysLysArgPheGlyLeuPheValGluAsnLeu 84
Db 1 AAAAGATACGATACCGTGTGATGAATGAAGCGTAGATTTAAGATCTTCTGAAATCTT 60
QY 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104
Db 61 CAACCTTATCAATCTACTAATAAGAAACCGCTCGGTATATCTCTGGTGTAAATCATTTT 120

```

```

QY 105 AlaAspLeuThrTyrGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
Db 121 GCTGATTGGACTTGGGAGGAGTTTCAAGAGTCATAGACTCGGTGCTGCTCAAAATGTTCT 180
QY 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
Db 181 GCTGCTCTCAAGGGAACCATAGGATTACCGATGTTTCTTCCCGCTGAGAAAGACTGG 240
QY 145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---TyrThr 163
Db 241 AGGAAAGAGAGGTATAGTCAGTGAAGTTAAGGATCAAGGCCACATCGGATCATGTGGACA 300
QY 164 PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
Db 301 TTCAGCAACACTGGAGCTTTGGAGTCAGCTTACGCCACAGCGCTTCGGAAGAATATCTCT 360
QY 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
Db 361 CTTTCTGAGCAGCAGCTAGTAGACTGTGCTGGTCTTACAATACTTTGGCTGCAATGGT 420
QY 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyIleAspThrGluGln 220
Db 421 GGGTTGCCATCCCAAGCCTTTGAATACATTAAATACAAATGCTGGCTTGAGACAGAGAA 480
QY 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
Db 481 GCATATCCCTACACTGGACAAATGGTCTCTGCAAAATTACATCTGAAAAGCTTGGCGTT 540
QY 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
Db 541 CAAGTCTCTGGCTCTGTCAATATCACCTGGGTGCTGAGGATGAATGAACATGCAAGTT 600
QY 261 GlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
Db 601 GCTTTTGTCTGGCCCGTTAGTGGCAATTCAGGTGGTGTGATGACTTCAGTTATACAAG 660
QY 281 LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
Db 661 AAAGGAGTTTACACTAGTACAACTTGTGGCAGCACACCCATGATGATGTAATCATCTGTT 720
QY 301 LeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIle 316
Db 721 CTTGCTGTGGTATGAATGAAGATGGTGATGATGATGATGATGATGATGATGATGATGAT 768
RESULT 14
LOCUS BE195255 794 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMR0088L13f Hordeum vulgare 5-45 DAP spike EST library
HVCNDA0009 (5 to 45 DAP) Hordeum vulgare subsp. vulgare cDNA clone
HVSMR0088L13f, mRNA sequence.
ACCESSION BE195255
VERSION BE195255.3 GI:16321152
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 794)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,
Fenton,R.D., Close,S.J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Morex 5-45 DAP spike cDNA library
JOURNAL Unpublished (2001)
COMMENT On Jun 26, 2000 this sequence version replaced gi:13188047.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu

```

Total hg bases = 385

Seq primer: AATTACCCCTCAATAAGG
High quality sequence stop: 489.

FEATURES

source
location/Qualifiers
1..794
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEH008813f"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library
HVCNA0009 (5 to 45 DAP)"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Plants were grown in the greenhouse at the University of
California, Riverside (Fenton, SJ Close, TJ Close). Whole
spikes with awns trimmed were collected at 5, 10, 15, 20,
30 and 45 DAP (Fenton). Total RNA was prepared from each
pool, equal quantities of all six RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Choi) in the TJ Close lab at the University of California,
Riverside. Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Xu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
this clone see <http://www.genome.clemson.edu/orders> Also
see Close TJ, Wing R, Kleinbofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

ORIGIN

Alignment Scores:
Pred. No.: 1,448-104 Length: 794
Score: 967.50 Matches: 183
Percent Similarity: 81.71% Conservative: 27
Best Local Similarity: 71.21% Mismatches: 40
Query Match: 51.88% Indels: 7
DB: 10 Gaps: 4

US-10-087-714-2 (1-352) x BE195255 (1-794)

Qy 99 LeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGly 118
Db 6 CTCGGCATCACCGTCTCGACATGAGCTGGAGAGATTCCAGCGCCGCTCTCGGC 65
Qy 119 AlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAsp 137
Db 66 GCGGCGCAGACCTGCTGGCGACGCTCGCCGCAACACCTCATCGGAGACGCCCGCG 125
Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValysAspGlnGly 157
Db 126 CTCCTGGAGACCAAGACATGGAGGAGGATGGATCGTCAGCCCGCTCAAAAACCGAGCC 185
Qy 158 SerCysGlySer---TrpThrPheSerThrGlyAlaLeuGluAlaAlaThrGln 176
Db 186 CATTGGCGTCTGCTGGACGTTGACACTACTGGCGCATTGAGCGACATATCTACG 245
Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
Db 246 GCCACTGGAAAGAACATCTCTCTTCTGAGCAACAGCTGGTTGACTGCTGGTGGGTTTC 305
Qy 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluThrValysTyra 213

Db 306 AATAACCTTTGGATGCAACGAGGCGCTTCCATCCAGCGCTTTCAGTACATCAATCAAT 365
Qy 214 GlyGlyIleAspThrGluGlnThrTrpTrpGlyValMetGlyIleCysAsnPhe 233
Db 366 GGGGGGATCGACACTGAGGAGTCTACCTCTTACAGGGTGTCAATGGCGTCTCGCATAC 425
Qy 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
Db 426 AAGCTGAATAATGCTCGGTTTCAGGTTTGGACTCAGTAACATCACACTGAATGCCGAG 485
Qy 254 AspGluLeuIysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 486 GATGAACGTGAAGACGCGCTCGGTTTGGTTCGCCAGTCAGGCGTGGCTTTCAGTGAT 545
Qy 274 LysGlyPheAsnLeuTyriLysGlyValTyriSerSerAspThrCysGlyArgAspPro 293
Db 546 GACGGGTTTACGAGTACACAGCCGAGTTTACCAACAGCCATTTGTGGCACTACCCCT 605
Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyriGlyValGluAspGlyIlePro-Ty 313
Db 606 GGTGACGTGAACCAACCGCTCTTGGCAGTCGGCTATGGTGTGAGAACGGGCGTCCGGTA 665
Qy 313 rTrpLeuIleIysAsnSerTrp-GlyThrAsnTrpGlyAspAsnGlyTyriPheIysMetG 333
Db 666 TTGGCTTATCAAGACTCGTGGGGCGCCGACTGGGGTGACAAATGGGTACTTCAAGATGA 725
Qy 333 lueuGlyIysAsnMetCysGlyValAlaAlaThrCysAlaSerTyriPro 348
Db 726 AATGGGAGAGACATGTGCTATTGGTAACCTGCGCCTTCTACCCCT 772

RESULT 15

BE131652

LOCUS

DEFINITION

L48-1651T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1651,
mRNA sequence.

ACCESSION

BE131652

VERSION

BE131652.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

720 bp mRNA linear EST 20-FEB-2001
An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
Unpublished (1997)
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L48-17
Seq primer: T3
High quality sequence stop: 350
POLYA=No.

FEATURES
source

Location/Qualifiers
1..720
/organism="Mesembryanthemum crystallinum"
/mol_type="mRNA"
/db_xref="taxon:3544"
/clone="L48-1651"
/tissue_type="Leaf, 48 h 0.4M NaCl"
/dev_stage="Six week old"
/clone_lib="Ice plant Lambda Uni-Zap XR expression

library, 48 hours NaCl treatment"
/notes=Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

ORIGIN

```
Alignment Scores:
Pred. No.:      2.24e-103      Length:      720
Score:          957.00         Matches:    179
Percent Similarity: 83.26%      Conservative: 20
Best Local Similarity: 74.90%    Mismatches: 36
Query Match:     51.31%         Indels:      4
DB:              10            Gaps:        3

US-10-087-714-2 (1-352) x BE131652 (1-720)

QY 115 AsnArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheVal 134
Db 3 CACAGATTGGAGCTGCCCAAACTGCTCTGCCACTCTCAAGGGCAGTCACAGCTGTCT 62
QY 135 AspGlyValLeuProValThrArgAspTTPArgGluGlnGlyLeuValSerProValLys 154
Db 63 CATGCAGACCTTCTCGAAACAAAAGATTGGAGAGCTGAGGGCATAGTCCGCCGTCGCAAG 122
QY 155 AspGlnGlySerCysGlySer---TTPThrPheSerThrThrGlyAlaLeuGluAlaAla 173
Db 123 AATCAACACCACTGTGGATCTCTGGACATTGAGTACAACTGGAGCTTTGGAGGAGCT 182
QY 174 TyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAla 191
Db 183 TATGGCGAGGCAATTTGGGAAGAACATCTCTTCTGAGCAACAACTTGTGATTGTGCT 242
QY 192 SerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVal 210
Db 243 GGTGCATTTGATTAATCATGTGTGCAATGGTGGTGGTTCATCCCAAGCTTTCGAGTATGC 302
QY 211 LysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIle 230
Db 303 AAGTACAATGAGGACTTGACACAGAGAGGCTTATCTTACACAGAAAGGATGGTGA 362
QY 231 CysAsnPheLysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeu 250
Db 363 TGTAAATCTTCAGCCCCAAAATGTTGGTGTTCAGTCTCTGGACTCCGTTAACATCACCCCTC 422
QY 251 GlyValaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPhe 270
Db 423 GAGCTGAAGAGAGACTTAAGGATGCTGTGGCATTTGGTCCGTCCTGTGAGTGTGGATT 482
QY 271 GluValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGly 290
Db 483 GAGGTTGTTAATGGATTCCGGTTTACACGAAAGAGATTACACACAGCACAACTTGTGGC 542
QY 291 ArgAspProMetAspValAsnHisAlaValAlaValGlyTyrGlyValGluAspGly 310
Db 543 AGCAGCCCCATGATGTAACCATCTGTTCTAGCTGTGGATGATGGAGTGGATGGT 602
QY 311 IleProTyrTrpLeuIleLysAsnSerTTPGlyThrAsnTrpGlyAspAsnGlyTyrPhe 330
Db 603 ATCCCATCTGGCTTATCAAGAACTCGTGGGAGAGATTGGGGTGACAAATGGATACTTC 662
QY 331 LysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
Db 663 AAGATGGAGATGGGAAAAAACAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 719
```

Search completed: March 30, 2004, 00:10:17
Job time : 2379 secs